

**Die Evaluation von Weiterbildungsprozessen in der Praxis:  
Status quo, Herausforderungen, Kompetenzanforderungen**

Eine Studie unter Organisationen, die das LQW-Modell anwenden

INAUGURAL-DISSERTATION  
zur Erlangung der Doktorwürde  
des  
Fachbereichs Erziehungswissenschaften  
der Philipps-Universität Marburg/Lahn

vorgelegt von  
Stefan Rädiker  
aus Bielefeld

Marburg/Lahn 2012

Vom Fachbereich Erziehungswissenschaften der Philipps-Universität Marburg  
als Dissertation angenommen am: 19.07.2012

Abschluss der mündlichen Prüfung am: 29.11.2012

Betreuer: Prof. Dr. Udo Kuckartz

Zweiter Gutachter: Prof. Dr. Rainer Zech

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis.....</b>	<b>6</b>
<b>Abkürzungsverzeichnis .....</b>	<b>7</b>
<b>Einleitung.....</b>	<b>9</b>
<b>1 Grundlagen der Evaluation .....</b>	<b>13</b>
1.1 Evaluation: ein strapazierter Begriff .....	13
1.2 Evaluation, Forschung, Evaluationsforschung .....	16
1.3 Die historische Entwicklung und der aktuelle Stand der Evaluation .....	19
1.4 Standards für Evaluationen.....	22
1.5 Evaluationszwecke: Die Frage nach dem Warum .....	24
1.6 Unterscheidungsmerkmale von Evaluationen .....	28
1.6.1 Wann wird evaluiert?	28
1.6.2 Wer evaluiert?	30
1.6.3 Wie unabhängig ist die Evaluation?	32
1.6.4 Wer erfährt von den Ergebnissen?	32
1.6.5 Welche Methoden und welche Designs kommen zum Einsatz?	33
1.7 Kompetenzanforderungen an Evaluierende .....	34
1.7.1 Internationale Kompetenzraster	38
1.7.2 Das Kompetenzraster der DeGEval	41
1.7.3 Konkretisierungen von Anforderungen an Evaluierende im Bildungsbereich	44
<b>2 Evaluation in der Weiterbildung .....</b>	<b>47</b>
2.1 Weiterbildung und Weiterbildungsprozesse: eine relevante Begriffsklärung .....	47
2.2 Evaluation von Weiterbildungsprozessen .....	52
2.3 Abgrenzung von Evaluation, Bildungscontrolling und Qualitätsmanagement.....	58
2.4 Unterscheidungsmerkmale für Evaluationen von Weiterbildungsprozessen .....	62
2.4.1 Zwecke: Warum wird evaluiert?	62
2.4.2 Gegenstände: Was wird evaluiert?	66
2.4.3 Fragestellungen: Mit welchem Fokus wird evaluiert?	68
2.4.4 Bewertungskriterien: Wie wird bewertet?	69
2.5 Methoden, Instrumente und Verfahren der Weiterbildungsévaluation.....	72
2.5.1 Befragung	73
2.5.2 Beobachtung	75
2.5.3 Verwendung bereits vorliegender Materialien	78
2.5.4 Tests und Lernkontrollen	78
2.5.5 Auswahl und Bewertung unterschiedlicher Evaluationsmethoden	80
2.6 Evaluationsansätze und -modelle mit Relevanz für die Weiterbildung .....	85
2.6.1 Der Vier-Ebenen-Ansatz von Kirkpatrick	87
2.6.2 Das CIPP-Modell von Stufflebeam	92
2.6.3 Der Programmbaum von Beywl et al.	96
2.6.4 Allgemeines Ablaufmodell von Evaluationen	100

2.7 Herausforderungen bei der Evaluation von Weiterbildungsprozessen .....	102
2.7.1 Widerstände gegen Evaluation .....	103
2.7.2 Wirkungsmessung und -nachweis .....	105
2.7.3 Unklare Gegenstandsziele .....	106
2.7.4 Auswahl und Definition von Indikatoren und Kriterien .....	106
2.7.5 Standardisierung und Vereinheitlichung von Instrumenten .....	107
2.7.6 Aufwands- und Ressourcenprobleme .....	108
2.7.7 Defizite bezüglich Nutzen und Nutzung .....	108
2.7.8 Fehlende Evaluationskompetenzen .....	108
<b>3 Die Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW)...</b>	<b>111</b>
3.1 Die Grundzüge des LQW-Modells.....	111
3.2 Die Evaluation von Bildungsprozessen im Rahmen des LQW-Modells.....	117
3.3 Forschungsergebnisse zur Evaluation von Weiterbildungsprozessen.....	119
<b>4 Methodisches Vorgehen bei der durchgeführten Studie.....</b>	<b>125</b>
4.1 Fragestellungen .....	125
4.2 Selbstbeschreibungen der Evaluationspraxis (Datenquelle I).....	126
4.2.1 Auswahl der Selbstreporte .....	128
4.2.2 Die qualitative Inhaltsanalyse .....	131
4.2.3 Maßnahmen zur Steigerung der Auswertungsqualität .....	136
4.3 Leitfadeninterviews mit Evaluationsverantwortlichen (Datenquelle II) .....	140
4.3.1 Der Leitfaden und die Interviewdurchführung .....	141
4.3.2 Der begleitende Online-Fragebogen .....	145
4.3.3 Das Interview-Sample .....	147
4.3.4 Transkription der Interviews .....	149
4.3.5 Die qualitative Inhaltsanalyse .....	150
4.4 Hinweise zum Umgang mit Quantifizierungen.....	151
4.5 Hinweise zur Verallgemeinerbarkeit der Ergebnisse.....	152
<b>5 Der Status quo der Evaluationspraxis.....</b>	<b>155</b>
5.1 Evaluationszwecke und Evaluationsgegenstände.....	155
5.2 Evaluationsverfahren, Evaluationsdesigns und Datenerhebung.....	162
5.3 Datenauswertung und Datenbewertung.....	169
5.4 Umgang mit Ergebnissen: Dokumentation, Kommunikation, Nutzung .....	171
5.5 Zusammenfassung und Diskussion des Status quo .....	175
5.6 Quintessenz für die Evaluationspraxis: Ein Fragenkatalog für die Reflexion und Gestaltung der Evaluation von Weiterbildungsprozessen.....	180
<b>6 Herausforderungen in der Evaluationspraxis .....</b>	<b>185</b>
6.1 Von Organisationen berichtete Herausforderungen in den Selbstreporten.....	185
6.2 Case Summaries vier ausgewählter Organisationen.....	187
6.2.1 Sozial/kulturell orientierte Organisation .....	188
6.2.2 Weltanschaulich gebundene Organisation .....	190
6.2.3 Volkshochschule .....	192
6.2.4 Kommerziell orientierte bzw. wirtschaftsnahe Organisation .....	194
6.3 Die Hauptproblembereiche .....	196
6.3.1 Evaluationsinstrumente und -verfahren .....	197
6.3.2 Datenerhebung .....	199

6.3.3 Akzeptanz und Belastung von Teilnehmenden, Lehrenden und Mitarbeitenden	200
6.3.4 Datenauswertung und -bewertung	203
6.3.5 Aufwand und Ressourcen	204
6.3.6 Nutzen, Erkenntnisgewinn und Aussagekraft	205
6.3.7 Nutzung, Umsetzung und Ergebnisverwendung	207
6.3.8 Weitere Einzelprobleme	207
6.4 Die sieben Hauptproblembereiche in der Gesamtschau	208
6.5 Entwicklungsperspektiven, Unterstützungsbedarf und Evaluationsverständnis	209
6.5.1 Geplante Änderungen der Evaluationspraxis	209
6.5.2 Unterstützungsbedarf bei der Evaluation	211
6.5.3 Gewünschte Lerninhalte eines Seminars zur Evaluation	212
6.5.4 Erhöhung des Nutzens	213
6.5.5 Verfolgte Evaluationszwecke	213
6.5.6 Evaluationsverständnis	214
6.6 Typologisierung der Organisationen hinsichtlich der Herausforderungen	215
6.6.1 Typ I: Instrumente und Verfahren im Fokus	218
6.6.2 Typ II: Ergebnisse im Fokus	219
6.6.3 Typ III: Aufwand im Fokus	220
6.6.4 Typ IV: Konzeption im Fokus	221
<b>7 Schlussfolgerungen: Ein Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen</b>	<b>223</b>
7.1 Von Organisationsvertretern/innen berichtete Kompetenzanforderungen	227
7.2 Das Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen im Detail	229
7.2.1 Basiswissen	230
7.2.2 Methodenkompetenzen	233
7.2.3 Organisationskenntnisse und pädagogische Kompetenzen	237
7.2.4 Lern- und Entwicklungskompetenzen	240
7.2.5 Nutzungskompetenzen	242
7.3 Übersicht über die Kompetenzbereiche des Anforderungsprofils	244
<b>8 Zusammenfassung und Ausblick</b>	<b>247</b>
<b>Danksagung</b>	<b>257</b>
<b>Literaturverzeichnis</b>	<b>259</b>
<b>Anhang</b>	<b>267</b>
A) Kategorienhandbuch für die Analyse der Selbstreporte	267
B) Leitfaden für die Interviews mit Evaluationsverantwortlichen	269
C) Anschreiben an die Interviewpersonen	270
D) Kategorienhandbuch für die Analyse der Interviews	271

# Abbildungsverzeichnis

Abb. 1:	Zusammenspiel von Evaluationszwecken (nach Stockmann, 2010c, S. 73) .....	26
Abb. 2:	Zusammenspiel von Programmphase, Analyseperspektive und Evaluationskonzepten (nach Stockmann, 2010c, S. 75) .....	29
Abb. 3:	Vor- und Nachteile interner und externer Evaluation (nach Stockmann, 2010c, S. 81) .....	31
Abb. 4:	Unterscheidungsmerkmale von Evaluationen (eigene Darstellung) .....	34
Abb. 5:	Auszug aus den Essential Competencies for Program Evaluators (ECPE) (Quelle: Stevahn et al., 2005, S. 49-51; Ghore et al., 2006, S. 110; Übersetzung nach Beywl, 2006, S. 333- 335 sowie eigene Übersetzung) .....	39
Abb. 6:	Auszug aus den Competencies for Canadian Evaluation Practice (CCEP) (Quelle: Canadian Evaluation Society, 2010, S. 5-9) .....	40
Abb. 7:	Auszug aus dem Rahmenwerk der European Evaluation Society (Quelle: European Evaluation Society, 2009, S. 21) .....	41
Abb. 8:	Kompetenzfelder und Dimensionen gemäß der DeGEval-Empfehlungen in Stichworten (Quelle: Caspari et al., 2008, S. 30-32) .....	42
Abb. 9:	Kompetenzanforderung an adult learning professionals bezüglich Evaluation (Quelle: Buiskool et al., 2010, S. 71) .....	45
Abb. 10:	Schritte des gemeinsamen Weiterbildungsprozesses von Lernenden und Organisation (eigene Darstellung) .....	49
Abb. 11:	Dimensionen des Weiterbildungsprozesses (eigene Darstellung) .....	51
Abb. 12:	Weiterbildungsprozess aus Sicht einer Organisation differenziert nach vier Dimensionen (eigene Darstellung) .....	51
Abb. 13:	Evaluation in der Weiterbildung auf verschiedenen Ebenen (eigene Darstellung) .....	53
Abb. 14:	Unterschiedliche Evaluationsgegenstände auf unterschiedlichen Evaluationsebenen (nach Gutknecht-Gmeiner, 2009, S. 13-6) .....	54
Abb. 15:	Kreislauf des Bildungscontrollings (Quelle: Markowitsch, 2009, nach Käßlinger, 2009, S. 2) .....	61
Abb. 16:	Das Zusammenspiel von Evaluationsgegenständen und Bewertungskriterien (eigene Darstellung) .....	71
Abb. 17:	Die vier Evaluationsebenen von Kirkpatrick (eigene Darstellung, mit Ergänzungen nach Reischmann, 2006, S. 111-122; Balzer, 2005, S. 54-55) .....	90
Abb. 18:	Der kausale Zusammenhang zwischen den vier Kirkpatrick-Ebenen (Quelle: Alliger & Janak, 1989, S. 335) .....	91
Abb. 19:	Integration des Vier-Ebenen-Ansatzes in das CIPP-Modell (eigene Darstellung) .....	96
Abb. 20:	Visuelle Darstellung des Programmbaums (Quelle: <a href="http://www.univation.org/download/Programmbaum_mit_Boxen.gif">http://www.univation.org/download/Programmbaum_mit_Boxen.gif</a> ) .....	97
Abb. 21:	Quellen von Widerständen in Lernprozessen und Evaluationsprozessen (nach Zuschlag, 1987, S. 78) .....	104
Abb. 22:	Der Lernerorientierte Qualitätstestierungsprozess (Quelle: Zech, 2006, S. 13) .....	113
Abb. 23:	Die Kontexte des Lernens (Quelle: ArtSet, 2006, S. 5) .....	116
Abb. 24:	Der fünfte Qualitätsbereich von LQW zur „Evaluation der Bildungsprozesse“ (Quelle: Zech, 2006, S. 68) .....	118
Abb. 25:	Zusammenhang zwischen Datenquellen und Forschungsfragen .....	126
Abb. 26:	Zusammensetzung der Zufallsstichprobe (n = 50) nach Organisationstyp, Organisationsgröße und Retestierung .....	130
Abb. 27:	Schematischer Ablauf der inhaltlich strukturierenden qualitativen Inhaltsanalyse (Quelle: Kuckartz, 2012, S. 78) .....	131
Abb. 28:	Hauptkategorien nach Abschluss der Analyse .....	133
Abb. 29:	Auszug der Subkategorien zur Hauptkategorie „Gegenstände/Inhalte/Fragestellungen“ .....	135
Abb. 30:	Beispiel für einige Zusammenfassungen von Textstellen der Subkategorie „Erfolgskontrolle, Zielüberprüfung“ der Hauptkategorie „Evaluationszweck“ .....	136
Abb. 31:	Überprüfung der Inter-coder-Übereinstimmung mithilfe unterschiedlicher Farben .....	137
Abb. 32:	Verschiedene Memosymbole für Notizen mit unterschiedlicher Funktion .....	140
Abb. 33:	Der begleitende Online-Fragebogen .....	146

Abb. 34: Anzahl angefragter und teilgenommener Organisationen.....	147
Abb. 35: Übersicht über das Interview-Sample.....	148
Abb. 36: Definitionen der Subkategorien zur Hauptkategorie „Problemfelder der Evaluation“ .....	151
Abb. 37: Identifizierte Evaluationsgegenstände und Inhalte auf den vier Dimensionen Lernen, Lehren, Lerninfrastruktur und Lernorganisation .....	158
Abb. 38: Unterschiede bezüglich Zwecken und Gegenständen der Evaluation bei verschiedenen Organisationstypen.....	161
Abb. 39: Unterschiede bezüglich der Evaluationsverfahren bei verschiedenen Organisationstypen ...	165
Abb. 40: Unterschiede bezüglich des Umfangs bei der Evaluation bei verschiedenen Organisationstypen und Organisationsgrößen .....	167
Abb. 41: Unterschiede bezüglich Konsequenzen und Nutzungsformen der Evaluationsergebnisse bei verschiedenen Organisationstypen .....	175
Abb. 42: Übersicht über die identifizierten Problemfelder (Quelle: Rädiker, 2012, S. 48).....	186
Abb. 43: Kausalkette von Instrument, Ergebnis und Nutzen .....	206
Abb. 44: Die sieben Hauptproblembereiche verortet im Ablauf eines Evaluationsprozesses .....	208
Abb. 45: Auszug aus einer Fallzusammenfassung für I09 .....	216
Abb. 46: Quellen für das Anforderungsprofil für Evaluierende von Weiterbildungsprozessen.....	223
Abb. 47: Auszug aus dem organisationsspezifischen Kompetenzmodell von ArtSet® (Quelle: ArtSet, 2012, S. 55).....	226
Abb. 48: Abgleich der DeGEval-Kompetenzfelder mit den Aussagen der Organisationsvertreter/innen.....	227
Abb. 49: Dimensionen des Kompetenzbereichs „Basiswissen“ .....	230
Abb. 50: Dimensionen des Kompetenzbereichs „Methodenkompetenzen“ .....	233
Abb. 51: Dimensionen des Kompetenzbereichs „Organisationskenntnisse und pädagogische Kompetenzen“ .....	238
Abb. 52: Dimensionen des Kompetenzbereichs „Lern- und Entwicklungskompetenzen“ .....	240
Abb. 53: Dimensionen des Kompetenzbereichs „Nutzungskompetenzen“ .....	243
Abb. 54: Kompetenzbereiche und Dimensionen des Anforderungsprofils in der Übersicht.....	244

## Abkürzungsverzeichnis

DgL	Definition gelungenen Lernens
HPM	hauptamtlich pädagogische Mitarbeitende
LQW	Lernerorientierte Qualitätstestierung in der Weiterbildung
MA	Mitarbeitende
päd. MA	pädagogische Mitarbeitende
TN	Teilnehmende





## Einleitung

Anfang der 1990er Jahre wurde in der Weiterbildung eine Debatte über das Thema Qualität geführt, die sich intensiv mit neuen Systemen des Qualitätsmanagements auseinandersetzte. Unter anderem wurde diskutiert, inwieweit sich in der Wirtschaft entwickelte Qualitätsmodelle wie ISO 9000 und Total Quality Management auf die Weiterbildung übertragen lassen und ob es überhaupt adäquate Qualitätsmodelle für die Weiterbildung geben könne (vgl. z.B. Arnold, 1994; Van den Berghe, 1998). Heute, gut 20 Jahre später, stellt sich diese Frage nicht mehr und sowohl die Beschäftigung mit Qualität als auch die Anwendung von Qualitätsmodellen in der Weiterbildung sind zur (verpflichtenden) Selbstverständlichkeit geworden. So schreiben einzelne Bundesländer öffentlich geförderten Weiterbildungsorganisationen seit mehreren Jahren vor, dass diese ein Qualitätsmanagementsystem einsetzen müssen, und es wurden zahlreiche Qualitätsmodelle für die Weiterbildung entwickelt (siehe die Übersicht bei Veltjens, 2006). Sehr stark verbreitet ist in Deutschland das Modell der „Lernerorientierten Qualitätstestierung in der Weiterbildung“ (LQW), das vor gut zehn Jahren von der Organisation ArtSet® mit Sitz in Hannover zusammen mit einem Netzwerk von Kooperationspartnern, unter anderem dem Deutschen Institut für Erwachsenenbildung, entwickelt wurde. Mehr als 550 Organisationen setzen heute in Deutschland, Österreich und einigen osteuropäischen Staaten das LQW-Modell ein. Unter ihnen findet sich ein breites Spektrum an Organisationen, das von Volkshochschulen über Unternehmen der beruflichen Bildung bis hin zu zahlreichen kulturellen, politischen, kirchlichen und anderen Bildungsträgern reicht. Weiterbildungsorganisationen, welche die Erfüllung von einheitlichen Mindestanforderungen in elf Qualitätsbereichen nachweisen können, erhalten ein bundesweit anerkanntes LQW-Testat.

Eng verwoben mit der gesamten Qualitätsdiskussion und folglich auch mit dem LQW-Modell ist der Begriff „Evaluation“, der in unterschiedlichen Variationen, Bedeutungen und Kontexten immer häufiger Verwendung im Bildungsbereich findet. Dabei wird jedoch auf unterschiedliche Ebenen Bezug genommen: Auf oberster Ebene wird die *Evaluation des Systems Weiterbildung* – meist von der Politik – angesprochen, worunter die Überprüfung der Weiterbildungslandschaft als Ganzes, losgelöst von einzelnen Organisationen, verstanden wird. Auf mittlerer Ebene sind beispielsweise die *Selbst- oder Fremdevaluationen von Organisationen* angesiedelt, mit denen in der Regel eine Mischung aus Legitimation, Organisationsentwicklung und Selbstvergewisserung verfolgt wird. Schließlich lässt sich auf unterster Ebene die

*Evaluation von Bildungsprozessen* verorten, welche sich vornehmlich auf die durchgeführte Bildungsarbeit richtet und im Mittelpunkt dieser Arbeit steht.

Bildungsprozesse zu evaluieren und sich über den Erfolg einer Bildungsveranstaltung zu vergewissern, ist im eigentlichen Sinne nichts Neues und zählt schon immer als wichtiger Bestandteil zum professionellen pädagogischen Handeln. So versteht Reischmann (2006, S. 86-89) Evaluation ausdrücklich als andragogisches Handeln und Will, Winteler und Krapp (1987b) stellen im Kontext der beruflichen Aus- und Weiterbildung fest: „Ausbilder und Dozenten, Bildungsmanagement und Unternehmensleitung waren schon immer daran interessiert, zielorientiert und systematisch Informationen über Bildungsmaßnahmen zu sammeln, zu sichten und zu bewerten“ (S. 35). Dieser prominenten Rolle von Evaluation trägt das LQW-Modell Rechnung, indem es einen eigenen Qualitätsbereich für die Evaluation von Bildungsprozessen beinhaltet, in dem mehrere von den Weiterbildungsorganisationen zu erfüllende Anforderungen definiert sind. Dadurch sind Organisationen, die LQW einsetzen, gezwungen, sich (zumeist systematischer und reflektierter als bisher) mit ihrer Evaluationspraxis, also unter anderem mit ihren eingesetzten Evaluationsinstrumenten und Verfahren und den evaluierten Gegenständen, auseinanderzusetzen und diese gegebenenfalls in Richtung der Mindestanforderungen von LQW zu verbessern bzw. zu verändern.

Es ist davon auszugehen, dass die Organisationen im Rahmen ihrer Evaluationspraxis auf Schwierigkeiten und Herausforderungen stoßen – zumindest die Umsetzung und Einführung von systematischer Evaluation, welche das LQW-Modell fordert, selbst als Herausforderung erleben. Diese Vermutung bestätigen mehrere Forschungsergebnisse, welche in dieser Arbeit im Detail vorgestellt werden. Beispielsweise berichten Hartz, Goeze und Schrader (2007, S. 135), dass in Organisationen, die das LQW-Modell anwenden, der Bereich Evaluation zusammen mit der Bedarfserschließung auf Platz 2 der wahrgenommenen Entwicklungsbedarfe steht. Unklar bleibt bei den bisher vorliegenden Forschungsergebnissen, um welche Entwicklungsbedarfe, Schwierigkeiten und Herausforderungen es sich konkret handelt. Dabei liegen viele Vermutungen nahe: Es könnten Widerstände bei Teilnehmenden, Lehrenden und Beschäftigten auftreten; es könnten sich zahlreiche Fragen zur methodischen Umsetzung stellen; es könnte sich ein geringer praktischer Nutzen der Evaluation herausstellen und die Organisationen könnten mit „Datenfriedhöfen“ zu kämpfen haben. Im Rahmen der vorliegenden Arbeit wurde eine Studie durchgeführt, welche klären sollte, inwieweit sich diese Vermutungen empirisch bestätigen

lassen und mit welchen Herausforderungen die Organisationen tatsächlich konfrontiert sind. Konkret berücksichtigt diese Arbeit bei der Untersuchung drei Schwerpunkte, die sich auch im Titel widerspiegeln: Erstens den Status quo der Evaluationspraxis, zweitens die Herausforderungen, welche bei der Evaluation der Bildungsprozesse auftreten können und drittens die Kompetenzanforderungen, die sich aus dem Status quo und den Herausforderungen ergeben.

Vor diesem Hintergrund gliedert sich die Arbeit wie folgt: Im ersten Kapitel wird der Blickwinkel zunächst weit gestellt und wichtige Grundlagen der Evaluation vorgestellt. Das zweite Kapitel fokussiert speziell die Evaluation in der Weiterbildung und geht auf relevante Spezifika dieses Feldes ein. Im dritten Kapitel wird das LQW-Modell in seinen Grundzügen vorgestellt und es werden die – oben bereits angesprochenen – relevanten und bisher vorliegenden Forschungsergebnisse präsentiert. Zu Beginn des vierten Kapitels werden zunächst die Forschungsfragen vorgestellt, welche die empirische Studie angeleitet haben, bevor anschließend das methodische Vorgehen im Detail erläutert wird. In Kapitel 5 werden die Studienergebnisse zum Status quo der Evaluationspraxis vorgestellt, welche insbesondere als Basis für die weiterführenden Analysen dienen, und in Kapitel 6 schließt sich die Präsentation der Ergebnisse zu den Herausforderungen an. Im siebten Kapitel werden die Untersuchungsergebnisse und die theoretischen Ausführungen unter der Perspektive von „Kompetenzanforderungen“ in Form eines Anforderungsprofils für Evaluierende in Weiterbildungsorganisationen zusammengeführt. Das achte und letzte Kapitel enthält eine Zusammenfassung der Ausführungen sowie einen Ausblick.

Die vorgelegte Arbeit ist aufgrund der behandelten Themen zu großen Teilen im Bereich der Forschung über Evaluation angesiedelt. Um die Arbeit innerhalb dieses breiten Feldes genauer zu lokalisieren, eignet sich ein Rückbezug auf die Arbeit von Mark und Henry (2004). Die beiden Autoren haben im Kontext der Forschung über Evaluationsnutzung eine sogenannte „Schematic Theory of Evaluation Influence“ vorgelegt, welche fünf verschiedene Bereiche unterscheidet: „Inputs der Evaluation“, „Aktivitäten der Evaluation“, „Outputs der Evaluation“ sowie „Mechanismen“ und „Outcomes der Evaluation“ (übersetzt und adaptiert von Hense, 2012). Mit Blick auf die jeweiligen Inhalte dieser einzelnen Felder lässt sich die hier vorgelegte Arbeit deutlich in den ersten beiden Feldern verorten. So greift die Analyse des Status quo der Evaluationspraxis die unterschiedlichen Phasen der „Aktivitäten der Evaluation“ auf, welche von der Ermittlung und Zusammenarbeit mit Beteiligten und Be-

troffenen bis hin zur Dissemination von Ergebnissen reichen. Die Analyse der Herausforderungen und die Erstellung eines Anforderungsprofils beziehen sich auf die „Inputs der Evaluation“, welche im Modell von Mark und Henry unter anderem auch die Kompetenzen und den fachlichen Hintergrund der Evaluierenden beinhalten.

Die Aktualität des behandelten Themas, insbesondere die Auseinandersetzung mit Kompetenzen von Evaluierenden in Weiterbildungsorganisationen, zeigt sich nicht zuletzt darin, dass seit März 2012 (vermutlich erstmalig) im deutschsprachigen Raum ein einjähriger berufsbegleitender universitärer Zertifikatslehrgang an der FH Nordwestschweiz zum Thema „Bildungsevaluation“ angeboten wird.<sup>1</sup> Der Lehrgang richtet sich an Fachpersonen aus dem Bildungsbereich, die sich für Datenerhebungen qualifizieren wollen. Die Teilnehmenden sollen erlernen, eine Vielzahl empirischer Erhebungsmethoden für Bildungsevaluationen maßschneidern und einsetzen zu können. In diesem Zusammenhang sei bereits angemerkt, dass im Mittelpunkt dieser Arbeit und der durchgeführten Studie zwar Organisationen stehen, welche das LQW-Modell anwenden. Wie im vierten Kapitel jedoch im Detail beschrieben wird, lassen sich die Ergebnisse der Studie mit wenigen Einschränkungen auch auf andere Weiterbildungsorganisationen übertragen.

---

1 Vgl. <http://www.fhnw.ch/ph/iwb/kader/bildungsevaluation>.

# 1 Grundlagen der Evaluation

Die ersten beiden Kapitel legen den theoretischen Grundstein für die durchgeführte Studie zur Evaluation von Weiterbildungsprozessen. Da die Diskussion um Evaluation in der Weiterbildung in großen Teilen an die allgemeine Diskussion über Evaluation anknüpft, wird der Fokus im Kapitel 1 zunächst weitgestellt und Evaluation im Allgemeinen beleuchtet, wobei auch Beispiele aus der Weiterbildung einfließen. Im weiteren Verlauf nimmt dann Kapitel 2 die Besonderheiten der Evaluation in der Weiterbildung und die für die Weiterbildung relevanten Modelle und Ansätze in den Blick. Dabei ist es manchmal hilfreich und notwendig, die Perspektive noch einmal für die allgemeine Evaluationsdiskussion zu öffnen, um sie dann wieder auf die Weiterbildung zu verengen.

## 1.1 Evaluation: ein strapazierter Begriff

„Der Begriff der Evaluation wird so häufig und beliebig gebraucht, daß Leser und Zuhörer nicht immer wissen, was gemeint ist.“ Diese Feststellung ist bereits über 25 Jahre alt und stammt aus dem früher weit verbreiteten Werk „Die Volkshochschule, Handbuch für die Praxis“ aus dem Jahre 1984.<sup>2</sup> Doch trotz ihres Alters hat diese Aussage auch heute noch Gültigkeit, vielmehr trifft sie nach zunehmender Verbreitung von Evaluationen in allen gesellschaftlichen Bereichen und fortgeschrittener Evaluationsdiskussion heute sicherlich noch mehr zu denn vor zwei Jahrzehnten. Die Gründe für die unklare Verwendung sind vielfältig. Zu den Hauptgründen zählt, dass in sehr vielen, breit gestreuten und sehr unterschiedlichen Anwendungsfeldern evaluiert wird und es kaum Disziplinen gibt, in denen keine Evaluation stattfindet. Die Bandbreite der derzeit 16 disziplin- bzw. themenbezogenen Arbeitskreise der „Gesellschaft für Evaluation“ (DeGEval) mag diese Vielfalt verdeutlichen; sie reicht von der beruflichen Bildung über die Entwicklungspolitik bis hin zum Umweltbereich (vgl. [www.degeval.de/arbeitskreise](http://www.degeval.de/arbeitskreise)). Hinzu kommt als zweiter wesentlicher Grund, dass Evaluation als Begriff nicht nur in der Wissenschaft und im Rahmen professionell durchgeführter Evaluation teilweise unterschiedlich und mit unterschiedlicher Schwerpunktsetzung verstanden wird, sondern auch in der Alltagssprache Einzug gehalten hat. Durch seine alltagssprachliche Verwendung verliert der Begriff jedoch „zunehmend seine wissenschaftliche Präzision“ (Stockmann,

---

<sup>2</sup> Vgl. Tietgens (1984).

2010a, S. 9). Vor diesem skizzierten Hintergrund erscheint es umso nötiger, den in dieser Arbeit verwendeten Evaluationsbegriff auszuweisen, zu konturieren und zu definieren.

Von der Wortbedeutung her, die meist im Lateinischen, aber auch im Englischen und Französischen gesucht wird, heißt evaluieren zunächst einmal *bewerten* (vgl. Balzer, 2005; Gollwitzer & Jäger, 2007; Stockmann, 2007; Stockmann & Meyer, 2010). Doch ist diese einfache Feststellung zu unpräzise, schließlich wird bei Evaluationen außerhalb des Alltags nicht irgendetwas von irgendjemandem nach irgendwelchen Kriterien bewertet (vgl. Kromrey, 2001a, S. 106). Die meisten Autoren/innen treffen daher in ihren Definitionen Aussagen über die evaluierten Gegenstände (was), die zu verwendenden Methoden (wie) und teilweise auch über den Sinn und Zweck einer Evaluation (warum) – und unterscheiden sich in ihrem Evaluationsverständnis hinsichtlich der inhaltlichen Ausfüllung genau dieser Bereiche.

Beywl, Kehr, Mäder und Niestroj (2007) definieren Evaluation in einer praxisorientierten Einführung wie folgt: „Evaluation ist eine spezifische wissenschaftsbasierte Dienstleistung, die insbesondere immaterielle Gegenstände (wie Bildung, soziale Arbeit, Öffentlichkeitsarbeit ...) unabhängig, systematisch und datengeschützt beschreibt und bewertet, sodass Beteiligte und Betroffene die erzeugten Ergebnisse für ausgewiesene Zwecke nutzen können“ (S. 5). Während diese Definition konkrete Gegenstände benennt und Hinweise für die Datenerhebung enthält, verbleibt die Aussage zu den Zwecken noch sehr allgemein. Mertens hingegen trifft keine konkreten Feststellungen zu den Gegenständen, schränkt jedoch den Zweck von Evaluationen stark auf die Entscheidungsfindung ein: „Evaluation is the systematic investigation of the merit or worth of an object (program) for the purpose of reducing uncertainty in decision making“ (Mertens, 1998, S. 219). Viele andere namhafte Autoren/innen zählen hingegen gleich mehrere Evaluationszwecke auf. So schreibt Patton (1997): „Program evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgements about the program, improve program effectiveness, and/or inform decisions about future programming“ (S. 23). Und im Glossar ihres Lehrbuches definieren Stufflebeam und Shinkfield (2007) Programmevaluation als „systematic collection, analysis, and reporting of descriptive and judgemental information about the merit and worth of a program’s goals, design, process, and outcomes to address improvement, accountability, and dissemination questions and increase understanding of the involved phenomena“ (S. 710).

In den letzten Definitionen taucht ein weiterer zu definierender, in der Evaluation häufig verwendeter Begriff auf, nämlich der des *Programms (program)*. Unter einem Programm wird üblicherweise ein Bündel an Maßnahmen verstanden, das aus einer Abfolge von Aktivitäten besteht, um bei einer Zielgruppe gewünschte Resultate zu erreichen (vgl. Beywl et al., 2007, S. 6; Beywl & Niestroj, 2009, S. 83). Nach Wittmann (1985, S. 23) ist ein Programm durch die Aspekte Programmträger/Personal, Handlungen, Klientel, Hilfsmittel, Ziele, Annahmen/Theorien und einer Unterstützungsfunktion gekennzeichnet. Aufgrund der historischen Entwicklungen, die in Abschnitt 1.2 beschrieben werden, nimmt die Programmevaluation einen sehr großen Stellenwert in der Diskussion um Evaluationen ein.<sup>3</sup> Da auch eine Weiterbildungsveranstaltung als Programm aufgefasst werden kann, konzentriert sich diese Arbeit ebenfalls auf die Programmevaluation, wenngleich natürlich auch andere Formen der Evaluation berücksichtigt werden.

Neben bekannten Autoren/innen der Evaluationsszene haben auch Fachvereinigungen für Evaluation ihr Evaluationsverständnis ausgewiesen. Die vorwiegend im deutschen Sprachraum aktive Gesellschaft für Evaluation (DeGEval) geht in ihrer Definition von Evaluation auf die Gegenstände und Methoden, nicht jedoch auf die Zwecke ein:

„Evaluation ist die systematische Untersuchung des Nutzens oder Wertes eines Gegenstandes. Solche Evaluationsgegenstände können z.B. Programme, Projekte, Produkte, Maßnahmen, Leistungen, Organisationen, Politik, Technologien oder Forschung sein. Die erzielten Ergebnisse, Schlussfolgerungen oder Empfehlungen müssen nachvollziehbar auf empirisch gewonnenen qualitativen und/oder quantitativen Daten beruhen.“ (DeGEval, 2008, S. 15)

Häufig wird in der Literatur die Wissenschaftlichkeit des eigenen Evaluationsverständnisses betont. Nach Stockmann (2007, S. 27; 2010c, S. 66) sind „wissenschaftliche“ Evaluationen im Gegensatz zu „Alltagsevaluationen“ durch folgende Punkte gekennzeichnet:

1. Sie beziehen sich auf einen klar definierten Gegenstand,
2. es werden objektivierende empirische Datenerhebungsmethoden eingesetzt,

---

3 Dabei wird die Evaluation von Programmen in der Literatur häufig gegen die Personal- und Produktevaluation abgegrenzt (vgl. etwa Beywl et al., 2007, S. 6).

3. es findet eine Bewertung anhand von Kriterien statt, die explizit auf den zu evaluierenden Sachverhalt bezogen sind sowie präzise festgelegt und offen gelegt wurden,
4. sie wird mit Hilfe systematischer vergleichender Verfahren vorgenommen,
5. sie wird in der Regel von „besonders befähigten“ Personen durchgeführt,
6. um Entscheidungen zu treffen, die auf den Evaluationsgegenstand bezogen sind.

Nimmt man die bisher vorgestellten Definitionen und die oft geforderte Wissenschaftlichkeit von Evaluationen als Maßstab, dann ist nicht jede Studie, die den Titel „Evaluation“ trägt, auch tatsächlich eine Evaluation. So weist Stockmann (2010c, S. 65-66) unter Bezugnahme auf Kromrey (2001a, S. 105-106) darauf hin, dass es sich nicht um Evaluation handele, wenn im Rahmen von Umfrageforschung nur die Befragten nach Bewertungen gefragt werden, im weiteren Verlauf der Evaluation aber keine weitere systematische Bewertung erfolge. Auch nicht jede Form eines Gutachtens stelle eine Evaluation dar.

Zusammenfassend lässt sich für diese Arbeit festhalten, dass Evaluation folgende Eckpunkte aufweist: Evaluation ist eine

- Beschreibung und kriteriengestützte Bewertung
- eines konkreten Gegenstandes
- anhand systematisch erhobener qualitativer und/oder quantitativer Daten.

Absichtlich wird an dieser Stelle keine Festlegung auf konkrete Gegenstände oder bestimmte Zwecke vorgenommen, da diese je nach Betrachtungsweise der verschiedenen Beteiligten einer Evaluation auch unterschiedlich ausfallen können (vgl. Abschnitt 1.6). Diese sehr offene Definition erscheint mit Blick auf die durchzuführende empirische Analyse angemessener zu sein als ein engeres Verständnis, da nicht abzusehen ist, welches Verständnis Weiterbildungsorganisationen vom Evaluationsbegriff haben und ein weiteres Begriffsverständnis sicherstellt, dass dieses bei der Analyse auch anwendbar ist.

## **1.2 Evaluation, Forschung, Evaluationsforschung**

Evaluation wendet zwar in der Regel sozialwissenschaftliche Methoden an, folgt in ihrem Ablauf, ihren Zielsetzungen und ihren Standards jedoch einer eigenen Logik. So liegt es nahe, nach Unterschieden zwischen einerseits Evaluation und anderer-



seits Forschung zu fragen. Für Glass besteht die zentrale Differenz darin, „dass pädagogische Evaluation den *Wert*, pädagogische Forschung dagegen die wissenschaftliche *Wahrheit* einer Sache einzuschätzen versucht“ (1972, S. 169; Hervorh. im Orig.). In den vergangenen Jahren sind die Unterschiede weiter herausgearbeitet worden und verschiedene Autoren/innen haben diese Differenzen zusammengetragen und teilweise auch systematisiert (unter anderem Balzer, 2005; Bortz & Döring, 2006; Kromrey, 2003; Reischmann, 2006; Stockmann, 2010c). Im Folgenden werden die Unterschiede bezüglich einiger wichtiger Kriterien dargestellt, wobei es zu berücksichtigen gilt, dass die Grenzen zunehmend verschwimmen, je weiter man sich von der Grundlagenforschung entfernt und der angewandten Forschung nähert.

*Zielsetzung:* Während Forschung vorrangig die allgemeine Wissensgewinnung, Theoriebildung und Theorieüberprüfung zum Ziel hat, steht bei Evaluationen meist die direkte Verbesserung von konkreter Praxis im Vordergrund, weshalb die Generalisierbarkeit von Ergebnissen nur eine untergeordnete Rolle spielt. Während eine Evaluation beispielsweise fragen würde, ob sich der höhere Druckpreis für einen Vierfarb-Druck des Bildungsprogramms lohnt, verfolgt Forschung eher die Frage, welche Farben bei Weiterbildungsprogrammen den besten Aufmerksamkeitswert haben (vgl. Reischmann, 2006, S. 28-29). Folglich ist für Evaluationen die Nützlichkeit und Verwertbarkeit der Ergebnisse für den evaluierten Gegenstand von besonders großer Bedeutung, was Beywl (2006, S. 116) in einer Fußnote überspitzt auf den Punkt bringt: „Forschung ohne Nutzung ist bedauerlich, Evaluation ohne Nutzung ist wertlos.“ Auf diesen häufig als entscheidend betrachteten Unterschied weist Glass bereits im Jahre 1972 hin: „Um Evaluatoren und Forscher unterscheiden zu können, empfiehlt sich zu fragen, ob man eine Untersuchung als Fehlschlag ansehen würde, wenn sie keine Informationen über den Nutzen des untersuchten Phänomens lieferte. Als Forscher wird man wahrscheinlich diese Frage verneinen“ (Glass, 1972, S. 170).

*Nutzung der Ergebnisse:* Die Ausrichtung der Evaluation hin auf Nutzung impliziert auch eine andere Bedeutung und einen anderen Umgang mit den Ergebnissen als dies bei der Forschung der Fall ist. Im Rahmen von Evaluationen wird häufig bereits zu Beginn die Ergebnissenutzung antizipiert, um den gesamten Evaluationsprozess auf die geplante Nutzung hin zu fokussieren. Diese Orientierung kommt in nutzungsorientierten Evaluationsansätzen wie etwa der „Utilization-focused Evaluation“ von Michael Quinn Patton, 1978 als Monographie in erster Auflage erschienen, zum Ausdruck. Dahingegen ist Forschung weniger auf die Nutzung ihrer Ergebnisse fokus-

siert, was natürlich nicht heißt, das daran kein Interesse bestünde. Die vollständige Nutzung der Ergebnisse ist jedoch kein Zwang und lässt sich auch nicht immer vorab planen und im Detail einschätzen.

*Stellenwert der Bewertung:* Jede Evaluation beinhaltet eine Bewertung, sie ist das typische Kennzeichen einer Evaluation. Demgegenüber kann sich Forschung „weitgehend normativer Urteile enthalten“ (Stockmann, 2004, S. 12) und unterliegt keinem Bewertungszwang.

*Fremdbestimmung:* Forschende können freier agieren als Evaluierende, denn die Auftraggebenden von Evaluationen bestimmen über bzw. nehmen starken Einfluss auf Fragestellung, Evaluationsgegenstand, Zeitrahmen und Finanzbudget. Der Kontext von Evaluationen ist häufig „politisch sensibel“ (Stockmann, 2010c, S. 59), da Evaluationen Entscheidungen vorbereiten und über den Fortbestand eines Programms entscheiden können. Forschende sind „zumindest vom Prinzip her nur sich, der scientific community und dem Erkenntnisgewinn verpflichtet, auch wenn die Praxis oft anders aussieht“ (Balzer, 2005, S. 19).

Eine Unterscheidung zwischen Evaluation und Forschung ist deutlich leichter vorzunehmen, als eine klare Grenze zwischen *Evaluation* und *Evaluationsforschung* zu ziehen, weil sich hierzu keine einheitliche Begriffsverwendung etabliert hat.<sup>4</sup> Anders als sich hinter Kindheitsforschung eindeutig die Forschung *über* die Kindheit verbirgt, verstehen nur wenige Autoren/innen unter Evaluationsforschung auch Forschung *über* Evaluation, wie dies etwa bei Beywl und Niestroj (2009, S. 38) der Fall ist.<sup>5</sup> In vielen Publikationen werden beide Begriffe, mal mit und mal ohne Hinweis, synonym verwendet (z.B. bei Stockmann & Meyer, 2010 oder bei Rossi, Freeman & Lipsey, 2006). Andere Autoren/innen ziehen den Begriff „Evaluationsforschung“ vor, um damit die Anwendung sozialwissenschaftlicher Methoden, die Einhaltung wissenschaftlicher Standards und die Verwendung expliziter Regeln bei der Bewertung zu betonen. Dementsprechend findet sich im Lehrbuch von Stufflebeam und Shinkfield (2007) im Glossar sowohl eine Definition für „Evaluation“ als auch für „Evaluation research“. Letztere fasst Evaluationsforschung als Typus sozialwissenschaftlicher Forschung auf. Lüders (2006, S. 46) bemerkt in diesem Zusammenhang,

---

4 Ursprünge dieser Unterscheidung werden teilweise bei Suchman (1967) verortet, der sich mit diesem Thema in seiner Monographie „Evaluative Research. Principles and practice in public service and social action programs“ auseinandersetzt und dort zwischen Evaluation und evaluierender Forschung unterscheidet.

5 Hirschauer (2006, S. 406) führt die in diesem Kontext sprachlich zu erwartende Begriffsbedeutung weiter aus.

„dass die allerorten beliebte Gegenüberstellung von Forschung und Evaluationsforschung auf – vorsichtig formuliert – wackeligen Prämissen beruht“ und verortet einem Vorschlag von Wensierski (2003) folgend Evaluation als intermediäres System zwischen professioneller Reflexion von Praxis einerseits und wissenschaftlich-analytischer Forschung andererseits. Zudem schlägt Lüders (2006, S. 49) vor, zwischen Evaluation im Sinne einer professionellen Praxis und Evaluationsforschung im Sinne eines besonderen Typs sozialwissenschaftlicher Forschung zu unterscheiden. Weil davon auszugehen ist, dass für Weiterbildungsorganisationen nicht primär der wissenschaftliche Anspruch, sondern die Nützlichkeit der durchgeführten Evaluationen im Vordergrund steht, sind die Evaluationen auch dort als Teil professioneller Praxis zu verstehen.

### **1.3 Die historische Entwicklung und der aktuelle Stand der Evaluation**

Dass Evaluation in der heutigen Zeit eine hohe Bedeutung für zahlreiche Bildungsfelder auf den unterschiedlichen Ebenen Forschung, Praxis und Politik innewohnt, ist unbestritten. Aber worauf ist diese hohe und vermutlich weiter steigende Relevanz zurückzuführen, wie ist es dazu gekommen und worin kommt sie zum Ausdruck?

Der Ursprung „moderner Evaluation“ (Rossi et al., 2006, S. 28; Stockmann, 2010b, S. 24), verstanden als ein systematisches und wissenschaftliche Regeln befolgendes Verfahren, wird meist in den USA der 1930er und 1940er Jahre verortet. Zu jener Zeit wurden zahlreiche Maßnahmen umfassend evaluiert, die im Rahmen des von Roosevelt eingeführten Programms „New Deal“ zur Reduzierung der Arbeitslosigkeit und Verbesserung der sozialen Sicherheit aufgelegt wurden. Die Entwicklung einer professionellen Evaluationskultur setzte sich in den USA in den 1950er und 1960er Jahren weiter fort, als sich zahlreiche weitere Programme im Sozial- und Bildungsbereich, aber auch im Infrastrukturbereich einer Wirksamkeitsprüfung stellen mussten. 1965 wurde ein Gesetz zur Erhöhung von Fördermitteln für den Bildungsbereich erlassen, deren Vergabe von der Durchführung einer Evaluation abhängig gemacht wurde. Besonders bekannt geworden ist das Projekt „Head Start“ (Laufzeit: 1965-1980), an dem 7,5 Millionen Kinder teilnahmen, zu dem über 50 Evaluationen durchgeführt und ca. 800 Artikel publiziert wurden (vgl. Gollwitzer & Jäger, 2007, S. 8; Stockmann, 2010b, S. 24-25). Das Instrument der Evaluation sollte also in der Regel sicherstellen und dazu beitragen, dass die mit viel Geld finanzierten Projekte und Programme zu einer tatsächlichen Verbesserung führten. Durch die

damals intensiv und zahlreich durchgeführten Evaluationen wurden diese „allmählich zu einem zentralen Bereich angewandter pädagogischer Forschung“ (Wulf, 1972, S. 9).

Die 1970er bis Mitte der 1980er Jahre in den USA werden häufig als „Zeitalter der Professionalisierung“ beschrieben, das gekennzeichnet war durch die Einrichtung von Studiengängen und Studienprogrammen, die Publikation von Fachliteratur und der Entwicklung von Evaluationsmodellen sowie der Gründung von Fachorganisationen und Netzwerken. So bildete sich 1976 das „Joint Committee on Standards for Educational Evaluation“, das 1981, also fünf Jahre später die „Standards for Evaluation of Educational Programs, Projects, and Materials“ veröffentlichte. 1986 wurde die bis heute sehr einflussreiche „American Evaluation Association“ gegründet, die 1989 erstmalig die „Program Evaluation Standards“ herausgab, die aus den zuvor genannten Standards hervorgegangen sind. Ende der 1980er Jahre gewinnen aufgrund von Kritik an sozialen Reformprogrammen unter Präsident Reagan Kosten-Nutzen-Fragen zunehmend an Bedeutung und Mitte der 1990er Jahre wurde für den Bundesetat die Input-Budgetierung durch die Output-Budgetierung ersetzt, das heißt, die Wirksamkeit von Politik und Maßnahmen rückte in den Vordergrund und Evaluation als angemessenes Mittel zu dessen Überprüfung ebenfalls. Bis heute ist die USA im Bereich der Evaluation hinsichtlich Diskussionsstand, institutioneller Verankerung und Evaluationskultur als führend anzusehen. Dies zeigt sich nicht zuletzt in zahlreichen Fachzeitschriften sowie weit entwickelten und zahlreichen Aus- und Weiterbildungsgängen und der Tatsache, dass die am meisten rezipierten Evaluationstheorien und -methoden US-amerikanischen Ursprung haben. (Vgl. Gollwitzer & Jäger, 2007, S. 9; Stockmann, 2010b, S. 26)

Die Entwicklung der Evaluation in Deutschland ist stark durch Nordamerika, aber auch die europäischen Entwicklungen geprägt, allerdings waren die ersten professionellen Bemühungen um Evaluation im Vergleich zu den USA erst deutlich später zu beobachten. Unter anderem ausgelöst durch die „Bildungskatastrophe“ in den 1960er Jahren wurden in Deutschland in den 1970ern zahlreiche Bildungsreformen angestoßen, die intensiv evaluiert wurden. Im Zentrum der Aufmerksamkeit stand in dieser Zeit die Curriculumevaluation in Schulen, zu dessen Praxis und Theorie Wulf im Jahre 1972 von ihm übersetzte US-amerikanische Grundlagentexte zur Evaluation von Cronbach, Scriven, Stake, Stufflebeam, Alkin, Glass und anderen unter dem Titel „Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen“ herausgab.

Nach einer Flaute in den 1980er Jahren hat die Evaluationsthematik in Deutschland, insbesondere angestoßen durch die allgemeine Qualitätsdebatte, seit Beginn der 1990er Jahre wieder Fahrt aufgenommen. „Hektische Hochkonjunktur“ betitelt Reichmann (2006, S. 6) diese Phase, die sich in einer stark zunehmenden Zahl an Publikationen niederschlägt, und mit einer großen Nachfrage nach Evaluationen und hohen, teilweise überzogenen Erwartungen gekennzeichnet ist.

In den letzten zehn Jahren hat das Thema Evaluation, insbesondere im Zusammenhang mit dem Thema Qualität, zunehmend an Normalität gewonnen. Diese Einschätzung stützt sich auf mehrere zu beobachtende Indikatoren, die eine zunehmende Professionalisierung der Evaluation in Deutschland belegen. Im Jahre 1997 wurde die Deutsche Gesellschaft für Evaluation (DeGEval, heute: Gesellschaft für Evaluation) gegründet, deren Mitgliederzahl in den letzten Jahren stark gestiegen ist und Mitte 2010 bei über 600 Mitgliedern lag, unter ihnen 113 Institutionen. Die DeGEval verfolgt als Hauptziele die Professionalisierung von Evaluation, die Zusammenführung unterschiedlicher Perspektiven und die Förderung von Information und Austausch. Mit diesen Zielen vor Augen verabschiedete die DeGEval auf ihrer Mitgliederversammlung im Jahre 2001 die „Standards für Evaluation“, die seit 2008 bereits in vierter, unveränderter Auflage vorliegen. Die Standards sollen zu einer einheitlichen Begriffsverwendung beitragen, die Qualität von Evaluationen erhöhen und sichern und sie bilden – wie die DeGEval selbst betont – den „Kern von mittlerweile mehreren Handreichungen“ ([www.degeval.de](http://www.degeval.de)). Allerdings muss man auch kritisch anmerken, dass es weiterer Anstrengungen bedarf, um die Anwendung und Verbreitung der Standards voranzutreiben (vgl. Balzer, 2005, S. 77; Brandt, 2009).

Zahlreiche weitere Indikatoren zeigen eine zunehmende Professionalisierung der Evaluation in Deutschland an (vgl. Stockmann, 2010b, S. 41-43). Es sind im letzten Jahrzehnt fortlaufend neue deutschsprachige Handbücher und Einführungsbücher zur Evaluation erschienen (z.B. Wottawa & Thierau, 2003; Gollwitzer & Jäger, 2007; Stockmann, 2006, 2007), ebenso zu speziellen Methodiken wie etwa der qualitativen Evaluation (z.B. Flick, 2006; Kuckartz, Dresing, Rädiker & Stefer, 2008). Im Jahre 1997 wurde die Mailingliste „Forum Evaluation“ mit heute über 400 Teilnehmenden eingerichtet und im Jahre 2009 entstand im Rahmen der 12. Jahrestagung der DeGEval die zunehmend aktive Mailingliste des Nachwuchses im Evaluationsbereich mit derzeit etwa 50 Mitgliedern. Und zuletzt, aber nicht minder bedeutsam, sei auf die seit 2002 erscheinende deutschsprachige „Zeitschrift für Evaluation“ mit ca. 600 Abonnenten und die Einrichtung und Etablierung von Master-Studiengängen im

deutschsprachigen Raum an der Universität des Saarlandes, der Universität Bern, der Universität Bonn und der Universität Krems hingewiesen.

Bezüglich eines eigenen und vor allem möglichst einheitlichen Evaluations-Vokabulars als Ausdruck einer Evaluationsprofession muss ergänzend noch das 2009 in zweiter Auflage erschienene Glossar der wirkungsorientierten Evaluation von Wolfgang Beywl und Melanie Nistroj erwähnt werden, das im Internet als Wiki für eine breite Öffentlichkeit zur Verfügung steht.<sup>6</sup> Dass es notwendig und sinnvoll ist, für eine eindeutige Begriffsverwendung innerhalb des Feldes Evaluation zu sorgen, zeigen auch die Bemühungen von OECD und EU, die im Internet mehrsprachige, synoptisch aufgebaute Richtlinien zum einheitlichen Verständnis von wichtigen Evaluationsbegriffen wie Impact, Outcome oder Results bereitstellen (vgl. EuropeAid, 2005).

Welches Fazit lässt der Blick auf die weltweite Entwicklung im Bereich Evaluation zu? Stockmann (2010b, S. 35) kommt zu dem Schluss, „dass sowohl für die meisten europäischen Länder als auch für die EU-Institutionen eine deutliche Zunahme an Politik- und Programmevaluationen zu verzeichnen ist, dass eine Vielzahl von Professionalisierungsbestrebungen zu erkennen sind und dass sich zunehmend eine ‚Evaluationskultur‘ ausbreitet“. Nach dem ersten Evaluationsboom der 1960er und 1970er sei deshalb in den aktuellen Entwicklungen ein zweiter, weitaus globaler angelegter Evaluationsboom zu sehen (ebd., S. 38).

## 1.4 Standards für Evaluationen

Auch wenn die im vorigen Abschnitt erwähnten Standards für Evaluation noch nicht die gewünschte flächendeckende Verbreitung und Anwendung gefunden haben, dienen sie zunehmend häufiger als Referenzpunkt von Evaluationen und werden in zahlreichen Handreichungen der DeGEval weiter konkretisiert, beispielsweise in den „Empfehlungen zur Didaktik der Evaluation“, die der Arbeitskreis Aus- und Weiterbildung<sup>7</sup> erarbeitet hat. Die Standards dienen nicht nur der Praxis, sondern auch der Auseinandersetzung mit Evaluation auf theoretischer und konzeptioneller Ebene, z.B. als Orientierungsrahmen in einem Sammelband von Kuckartz und Rädiker (2012). Zu ihrer allgemeinen Funktion ist zu bemerken, „dass die Evaluationsstandards eine die Evaluationspraxis unterstützende, von verschiedenen theoretischen Ansätzen unabhängige Hilfe sind, die die Evaluationsqualität präzise beschreiben

---

6 Zu finden unter [www.eval-wiki.org/glossar/Eval-Wiki:\\_Glossar\\_der\\_Evaluation](http://www.eval-wiki.org/glossar/Eval-Wiki:_Glossar_der_Evaluation).

7 Vgl. Gutknecht-Gmeiner, Hense, Kihm, Lüth und Steckhan (2011)

und der Vielfalt der Evaluationslandschaft gerecht werden“ (Balzer, 2005, S. 77-78).<sup>8</sup> Es erscheint daher sinnvoll und zweckmäßig die Standards an dieser Stelle vorzustellen, um sie auch als Referenzpunkt für diese Arbeit verwenden zu können, zumal sie ergänzend zu Abschnitt 1.1 zur Begriffsklärung beitragen und auch auf Problem- punkte in der Durchführungspraxis hindeuten, da sie aufgrund ihrer Anlage für zahlreiche kritische Bereiche der Evaluation Richtlinien und Maßstäbe formulieren.

Bei den Evaluationsstandards handelt es sich nicht um Minimalstandards, die in jedem Fall erfüllt sein müssen, sondern um Maximalstandards, die bestmöglich eingehalten werden sollen. Dieses Verständnis ist wichtig zu erwähnen, da nicht in allen durchgeführten Evaluationen auch alle Standards anwendbar sind. Insgesamt liegen 25 Evaluationsstandards vor, die sich auf die vier Bereiche Nützlichkeit, Durchführ- barkeit, Fairness und Genauigkeit aufteilen, da Evaluationen nach dem Verständnis der DeGEval<sup>9</sup> diese vier grundlegenden Eigenschaften aufweisen sollten (vgl. DeGEval, 2008). Zu jedem der vier Bereiche sind im Folgenden die jeweiligen Erläu- terungen sowie exemplarisch der jeweils erste Standard ausgeführt; die weiteren Standards sind auf der Internetseite der DeGEval einsehbar ([www.degeval.de](http://www.degeval.de)):

„Die *Nützlichkeitsstandards* sollen sicherstellen, dass die Evaluation sich an den ge- klärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.

*Standard N1: Identifizierung der Beteiligten und Betroffenen* Die am Evalua- tionsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Perso- nengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.

...

---

8 Weitere Funktionen von Standards zur Programmevaluation listet u.a. Widmer (2000, S. 85-86) auf: Handlungsanleitung bei der Planung und Ausführung von Evaluationen; didaktisches Instrument in Aus- und Weiterbildung für (angehende) Evaluatoreninnen; Instrument zur systematischen Selbstre- flexion im Evaluationsprozess; Bewertungsgrundlage in Meta-Evaluationen; Bestandteil von Eva- luationskontrakten, Reglementen, Weisungen, Handbüchern u.ä.; identitätsstiftende Grundlage der Evaluationsgemeinde.

9 Diese Einteilung und dieses Verständnis lehnt sich ausdrücklich an die Standards des Joint Committee on Standards for Educational Evaluation (JC-Standards 2000/1994) an, das ursprüng- lich ebenfalls vier Bereiche unterscheidet: Utility, Feasibility, Propriety, Accuracy (vgl. [www.jcsee.org](http://www.jcsee.org)). Im Jahre 2010 kam ein fünfter Bereich zu den Program Evaluation Standards hin- zu, der mit „Evaluation Accountability“ überschrieben ist.

Die *Durchführbarkeitsstandards* sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.

*Standard D1: Angemessene Verfahren* Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.

...

Die *Fairnessstandards* sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.

*Standard F1: Formale Vereinbarungen* Die Pflichten der Vertragsparteien einer Evaluation (was, wie, von wem, wann getan werden soll) sollen schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese neu auszuhandeln.

...

Die *Genauigkeitsstandards* sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

*Standard G1: Beschreibung des Evaluationsgegenstandes* Der Evaluationsgegenstand soll klar und genau beschrieben und dokumentiert werden, so dass er eindeutig identifiziert werden kann.

...“ (DeGEval, 2008, S. 10-13; teilweise kursive Hervorhebungen ergänzt bzw. fette Auszeichnungen durch kursive ersetzt)

Die Anzahl der Standards innerhalb der Themenbereiche variiert und deutet eine Schwerpunktsetzung an. Während die Genauigkeit und die Nützlichkeit mit neun bzw. acht Standards unterlegt sind, werden Fairness und Durchführbarkeit nur mit drei bzw. fünf Standards konkretisiert.

## 1.5 Evaluationszwecke: Die Frage nach dem Warum

Die bisherigen Ausführungen, unter anderem zu den Nützlichkeitsstandards, sollten deutlich gemacht haben, dass die Durchführung einer Evaluation nicht zweckfrei geschieht, sondern Evaluationen einen Nutzen mit sich bringen sollen oder besser ge-



sagt: müssen. Dieser Nutzen kann unterschiedlicher Art sein, weshalb in der Literatur verschiedene Evaluationszwecke unterschieden werden. Statt vom Evaluationszweck wird dabei häufig auch von Evaluationsfunktionen, Evaluationszielen oder von den Aufgaben einer Evaluation gesprochen. Gemeinsam ist diesen verschiedenen Bezeichnungen, dass es immer um die Frage geht, wozu, weshalb oder warum, also zu welchem Zweck eine Evaluation durchgeführt wird und „was die Evaluation in Bezug auf den Evaluationsgegenstand und seine veränderbaren Bedingungen bewirken soll“ (Beywl & Niestroj, 2009, S. 43).

Um sprachliche Klarheit zu gewährleisten, ist es wichtig, zwischen den Zielen der Evaluation und den Zielen, die mit einem evaluierten Gegenstand verfolgt werden, zu unterscheiden. Während das Ziel des Gegenstandes „Bildungsmaßnahme“ die Integration von Arbeitslosen sein kann, könnte zugleich das Ziel der Evaluation dieser Bildungsmaßnahme darin liegen, deren Effektivität nachzuweisen. Um diese Unterscheidung begrifflich zu erleichtern, spricht man üblicherweise von Gegenstandszielen und Evaluationszwecken (vgl. DeGEval, 2008, S. 15).

Zwar ließen sich fast beliebig viele verschiedene Evaluationszwecke benennen, allerdings kommen einige Zwecke besonders häufig vor bzw. lässt sich der Großteil von ihnen einem der folgenden vier Zwecke zuordnen:

1. Wissens- und Erkenntnisgenerierung
2. Entscheidungsfindung
3. Verbesserung, Optimierung, Entwicklung
4. Rechenschaftslegung, Legitimation

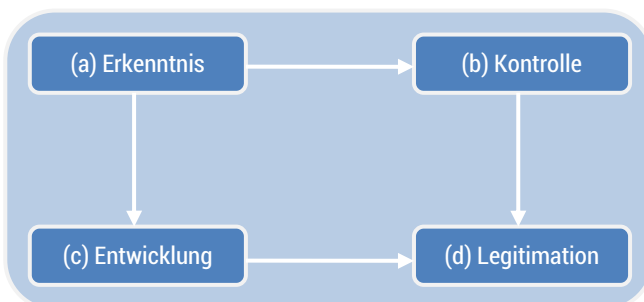
Einige Autoren/innen unterscheiden lediglich zwischen den ersten oder letzten drei Zwecken, weisen dafür aber zusätzlich auf sogenannte verdeckte Zwecke oder auf einen Kontrollzweck hin (Balzer, 2005, S. 199-201; z.B. Kromrey, 2001a, S. 114-115). Patton (2008, S. 139) führt in der vierten Auflage seines Lehrbuches „Utilization-Focused Evaluation“ aus Nutzungsperspektive sogar sechs verschiedene „Evaluation Purposes“ an: Overall Summative Judgement, Formative Improvement and Learning, Accountability, Monitoring, Developmental und Knowledge Generating. Gleich ob man drei, vier oder sechs Hauptzwecke unterscheidet, hat die Differenzierung eher analytischen Charakter, denn in der Praxis verschwimmen die Grenzen zwischen verschiedenen Evaluationszwecken schnell. Dies ist unter anderem darauf zurückzuführen, dass im Rahmen einer Evaluation meist mehrere Zwecke gleichzeitig im Fokus stehen und zugleich, dass unterschiedliche Stakeholder ei-

ner Evaluation häufig auch unterschiedliche Zwecke mit der Evaluation verbinden und verfolgen. Während beispielsweise für das pädagogisch-planende Personal in einer Bildungsorganisation die Entscheidung über die Einstellung oder Fortführung eines Kurses im Vordergrund steht, kann für die Lehrenden vielmehr die Verbesserung der Kurskonzeption durch die Evaluation von Bedeutung sein.

Gerade weil unterschiedliche Stakeholder unterschiedliche Interessen mit einer Evaluation verfolgen können, in der Praxis aber eine Evaluation nicht allen Zwecken gleichsam gerecht werden kann, wird oftmals darauf hingewiesen, dass für die Durchführung eines Evaluationsprojekts zu klären ist, welcher Zweck Priorität hat (vgl. Beywl & Balzer, 2008, S. 8). Der Evaluationszweck hat nämlich eine große Bedeutung für den Ablauf einer Evaluation: „Der Evaluationszweck richtet die Evaluation aus und gibt eine Orientierung, wozu die Evaluation als Prozess und wozu ihre Ergebnisse genutzt werden sollen“ (Beywl et al., 2007, S. 14). Konkret bestimmt die Zielsetzung der Evaluation beispielsweise darüber, welche Personen mit einbezogen werden sollten. Dieser Zusammenhang drückt sich auch im Nützlichkeitsstandard N2 „Klärung der Evaluationszwecke“ aus, in dem es heißt: „Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, so dass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann“ (DeGEval, 2008, S. 10).

Stockmann (2007, S. 37; 2010c, S. 73) verdeutlicht das Zusammenspiel verschiedener Evaluationszwecke anhand einer Grafik, die in Abb. 1 in vereinfachter Form wiedergegeben ist.

**Abb. 1: Zusammenspiel von Evaluationszwecken (nach Stockmann, 2010c, S. 73)**



Jede Evaluation liefert (a) Wissen und Erkenntnisse, z.B. um rationale Entscheidungen treffen zu können: Hat der Programmablauf reibungslos funktioniert? Haben die Maßnahmen die Zielgruppen erreicht? Wie groß ist die Akzeptanz des Programms? Welche Beiträge hat das Programm zur Zielerreichung beigetragen? Wie nachhaltig ist das Programm? Bei der (b) Ausübung von Kontrolle steht nicht die Entschei-

dungsfindung im Vordergrund, sondern die Evaluation dient insbesondere der Überprüfung, ob die anvisierten Ziele erreicht wurden. Mit jeder Evaluation ist direkt oder indirekt eine Form von Kontrolle verbunden, weil Evaluationen beispielsweise offen legen, ob alle an einem Programm Beteiligten ihre Aufgaben erfüllen, ihre Kompetenzen ausreichen etc. Die Befunde aus erkenntnis- und kontrollorientierten Evaluationen können genutzt werden, um das Programm zu optimieren und weiterzuentwickeln. In diesem Fall stehen (c) Entwicklungs- und Lernprozesse im Vordergrund. Die im Rahmen einer Evaluation gewonnenen Erkenntnisse und Ergebnisse einer Kontrollüberprüfung können gegebenenfalls zusammen mit Belegen der durchlaufenen Entwicklung dazu dienen, die durchgeführte Arbeit (d) zu legitimieren. Programmverantwortliche können mithilfe der Evaluation nachweisen, welcher Output mit welchen Vorgehensweisen und mit welchem Input erreicht wurde. In Zeiten von Finanzkrisen und knapper werdenden öffentlichen Kassen nimmt dieser Zweck eine zunehmend prominente Rolle ein. (Vgl. Stockmann, 2010c, S. 73-74)

Wie man in der Beratungsarbeit von „vergifteten Aufträgen“ spricht (z.B. Grochoviak & Castella, 2001, S. 237-239), wenn eigentlich gar keine Beratung gewünscht ist, sondern vielleicht nur die Durchsetzung von unliebsamen Entscheidungen durch die Beratenden, so sind auch im Bereich der Evaluation „vergiftete Aufträge“ anzutreffen. Stockmann (2010c, S. 75) nennt diese Evaluationsfunktion „taktische Funktion“, wenn Evaluationen lediglich bereits getroffene Entscheidungen (teilweise nachträglich) legitimieren sollen. Dies lasse sich jedoch nicht mit dem eigentlichen Zweck von Evaluationen vereinbaren und stelle eher ihre „pathologische Seite“ dar.<sup>10</sup> Stufflebeam und Shinkfield (2007) nennen derartige Evaluationen „Pseudoevaluationen“ und definieren sie wie folgt in ihrem Glossar:

**“Pseudoevaluations** Purposely produce and report invalid assessments or withhold or selectively release findings to right-to-know audiences. They do not adhere to professional standards for evaluation. Five predominant types are public relations, politically controlled, pandering, evaluation by pretext, and empowerment under the guise of evaluation.” (S. 710)

---

10 Für den Bereich der Weiterbildung berichtet Gieseke (2002) in einem Essay über ihre Erfahrungen mit dieser Schattenseite der Evaluation.

## 1.6 Unterscheidungsmerkmale von Evaluationen

Evaluationen lassen sich nicht nur bezüglich verschiedener Zwecke unterscheiden. In der Evaluationsliteratur finden sich darüber hinaus zahlreiche Merkmale von Evaluationen, anhand derer sich diese charakterisieren, typisieren und systematisieren lassen.<sup>11</sup> Die Unterscheidungsmerkmale können im Wesentlichen zwei Funktionen übernehmen. Erstens haben sie aus abstrakter, wissenschaftlicher Sicht einen ordnenden, klassifizierenden und analytischen Nutzen und werden deshalb auch im Rahmen von Evaluationsmodellen und Evaluationsansätzen verwendet bzw. wurden bei deren Entwicklung überhaupt erst eingeführt (vgl. Abschnitt 2.6). Zweitens können die Unterscheidungsmerkmale der Evaluationspraxis insbesondere während der Konzeptionsphase, aber auch während der Evaluationsdurchführung zur Reflexion dienen. Die Unterscheidungsmerkmale sind nicht immer trennscharf voneinander abgrenzbar, weil nicht alle Merkmale auf dem gleichen Abstraktionsniveau liegen bzw. weil sie ganz unterschiedliche Aspekte in den Blick nehmen.

Die folgende Darstellung der verschiedenen Unterscheidungsmerkmale wurde aus verschiedenen Publikationen zusammengetragen, und zwar aus Lehrbüchern (u.a. Gollwitzer & Jäger, 2007; Stockmann & Meyer, 2010; Wottawa & Thierau, 2003), aus richtungsweisenden Zeitschriftenartikeln (u.a. Balzer, Frey & Nenniger, 1999) sowie einschlägigen Evaluationslexika (u.a. Scriven, 1991; Beywl & Niestroj, 2009). Eine Darstellung von Unterscheidungsmerkmalen kann niemals vollständig sein; die hier vorgestellten Merkmale wurden mit Hinblick auf die spätere Analyse der Evaluationspraxis ausgewählt und es werden vorwiegend diejenigen vorgestellt, die vermutlich in der Evaluation von Weiterbildungsprozessen von Bedeutung sind. Sofern Begrifflichkeiten unterschiedlich verwendet werden (und mir dies bekannt war), wird darauf hingewiesen.

### 1.6.1 Wann wird evaluiert?

Eine der bekanntesten und am häufigsten verwendeten Unterscheidungsmerkmale ist die von Scriven 1972 eingeführte Trennung von *formativer* und *summativer* Evaluation. Eine formative Evaluation ist prozessbegleitend ausgerichtet und hat das Ziel, während der Durchführungsphase Einfluss auf das evaluierte Programm aus-

---

11 Diese Unterscheidungsmerkmale werden in der Literatur unterschiedlich betitelt. So liest man unter anderem „Arten von Evaluationsforschung“ (Gollwitzer & Jäger, 2007, S. 11), „Evaluationsformen“ (Holla, 2002, S. 110-111) und „Evaluationskonzeptionen“ (Wesseler, 2010, S. 1041), während beispielsweise Reischmann (2006, S. 105) von „Evaluationstypen“ spricht.

zuüben und dieses zu verändern. Formative Evaluationen werden daher vorwiegend durchgeführt, wenn als Evaluationszweck die Optimierung des Evaluationsgegenstandes im Vordergrund steht. Dahingegen soll die summative Evaluation „vor allem bilanzierende Schlussfolgerungen über die Güte und/oder Tauglichkeit eines Evaluationsgegenstands“ (Beywl & Niestroj, 2009, S. 97) hervorbringen.

Stockmann (2010c, S. 75) unterscheidet neben formativ und summativ in Rückgriff auf Scriven (1991, S. 169) auch noch *preformative Evaluationen* und erläutert die Bedeutung dieser drei Evaluationskonzepte anhand des Zusammenspiels von Evaluationszweck, Programmphase und Analyseperspektive, das er in einer tabellari-schen Übersicht darstellt. Abb. 2 gibt diese Übersicht in gekürzter Form wieder.

**Abb. 2: Zusammenspiel von Programmphase, Analyseperspektive und Evaluationskonzepten (nach Stockmann, 2010c, S. 75)**

Programmphase	Analyseperspektive	Evaluationskonzepte
Planung und Entwicklung	ex-ante	preformativ, teilweise auch formativ: aktiv gestaltend, prozessorientiert, konstruktiv
Durchführung und Implementation	on-going	formativ, teilweise auch summativ: beides möglich
Wirkung	ex-post	summativ: zusammenfassend, bilanzierend, ergebnisorientiert

Während der Planung und Entwicklung eines Programms findet die Bewertung meist ex-ante (aus vorausschauender Sicht) statt und dementsprechend dienen *pre-formative Evaluationen* dazu, bereits in der Planungsphase vor der Durchführung eines Programms mögliche negative Verläufe und Auswirkungen zu entdecken und zu verhindern und einen möglichst positiven Verlauf zu ermöglichen. Während der Programmdurchführung dienen *formative, on-going Evaluationen* dazu, die Steuerung des Programms zu unterstützen, insbesondere, um Entscheidungshilfen bereitzustellen und – einem Regelkreis ähnlich – die Zielerreichung und die Umsetzung des geplanten Programms zu überwachen. Nach Abschluss eines Programms, auch unmittelbar nach dessen Ende, dienen *summative Evaluationen* unter einer ex-post Perspektive (aus rückblickender Sicht) insbesondere dazu, die Wirkungen und die Nachhaltigkeit des Programms zu bewerten und die Kausalitätsfrage zu beantworten, auf welche Ursachen die Wirkungen zurückzuführen sind. (Vgl. Stockmann, 2010c, S. 76)

Die Unterscheidung in formative und summative Evaluation zieht auch Konsequenzen bezüglich der Rolle der Evaluierenden nach sich, auf die Kuper (2005, S. 18) hinweist: „Die Rolle des Evaluators in der formativen Evaluation ist die eines Bera-

ters bei der Gestaltung und Durchführung eines Programms; die Rolle des Evaluators in der summativen Evaluation ist die eines Urteilenden, der ein Programm abschließend bewertet.“

Wottawa und Thierau (1998, S. 33-34) grenzen bezüglich Evaluationen, die im Vorhinein Aussagen über den Erfolg von Maßnahmen treffen sollen, die *prognostische* und die *prospektive Evaluation* voneinander ab. Von prognostischer Evaluation wird gesprochen, wenn anhand der Bewertung eines real durchgeführten Programms der Erfolg weiterer gleicher Programme vorhergesagt wird. Dies wäre der Fall, wenn eine Bildungseinrichtung an mehreren Standorten ein neues Blended-Learning-Kursformat einführen möchte und zunächst an einem Standort den Erfolg und die Akzeptanz bei ihren Lernenden evaluiert. Wird ein Programm vor seiner Implementierung bewertet, und werden dabei nur die Voraussetzungen und Rahmenbedingungen analysiert und keine exemplarischen oder stichprobenhaften Programmumsetzungen berücksichtigt, spricht man von prospektiver Evaluation. (Vgl. auch Gollwitzer & Jäger, 2007, S. 14-15)

### 1.6.2 Wer evaluiert?

Bezüglich der Frage, wer die Evaluation durchführt, existieren mehrere wichtige zu treffende Unterscheidungen. Während bei der *Selbstevaluation* die Evaluierenden zu dem Personenkreis gehören, der für den Evaluationsgegenstand verantwortlich ist, werden *Fremdevaluationen* von Personen durchgeführt, die nicht am Gegenstand beteiligt sind.<sup>12</sup> Häufig wird synonym für Fremdevaluation auch der Begriff *externe Evaluation* und für Selbstevaluationen der Begriff *interne Evaluationen* verwendet. Jedoch differenziert Stockmann diesbezüglich zwischen der Organisation, in der ein zu evaluierendes Programm angesiedelt ist, und den Personen, die das Programm auf operativer Ebene durchführen. Nach Stockmann (2010c, S. 80) liegt eine interne Evaluation vor, wenn die Evaluation von der gleichen Organisation vorgenommen wird, die auch das zu evaluierende Programm durchführt. Stockmann spricht erst von Selbstevaluation, wenn die Evaluation auch von den Personen vorgenommen wird, die direkt an der Durchführung des evaluierten Programms beteiligt sind. Wenn also die zentrale Stelle für Lehrevaluation einer Universität eine Lehrveranstaltung eines Professors evaluiert, würde man nach Stockmann von interner, aber

---

12 Entsprechend definiert die Gesellschaft für Evaluation Selbstevaluation wie folgt: „Unter Selbstevaluation werden systematische, datenbasierte Verfahren der Beschreibung und Bewertung verstanden, bei denen die praxisgestaltenden Akteure identisch sind mit den evaluierenden Akteuren“ (DeGEval, 2004, S. 5).

nicht von Selbstevaluation sprechen.<sup>13</sup> Externe Evaluationen werden laut Stockmann von Personen durchgeführt, die weder dem Mittelgeber noch der Durchführungsorganisation angehören.

Sowohl mit interner als auch mit externer Evaluation sind in der Regel Vor- und Nachteile verbunden, von denen die wichtigsten in Abb. 3 festgehalten sind.

**Abb. 3: Vor- und Nachteile interner und externer Evaluation (nach Stockmann, 2010c, S. 81)**

intern	extern
<b>Vorteile:</b> <ul style="list-style-type: none"> <li>• rasch, geringer Aufwand</li> <li>• hohe Sachkenntnis</li> <li>• unmittelbare Umsetzung</li> </ul>	<b>Vorteile:</b> <ul style="list-style-type: none"> <li>• große Methodenkompetenz</li> <li>• hohe Unabhängigkeit und Distanz</li> <li>• große Glaubwürdigkeit</li> </ul>
<b>Nachteile:</b> <ul style="list-style-type: none"> <li>• geringe Methodenkompetenz</li> <li>• fehlende Unabhängigkeit und Distanz</li> <li>• „Betriebsblindheit“</li> </ul>	<b>Nachteile:</b> <ul style="list-style-type: none"> <li>• geringe Sachkenntnis</li> <li>• Abwehrreaktionen</li> <li>• Umsetzungsprobleme</li> </ul>

Zu den Vorteilen der internen Evaluation zählt die meist unaufwändige Durchführung sowie die hohe Sachkenntnis der Mitarbeitenden einer Organisation, aber auch die Möglichkeit der direkten Ergebnisumsetzung. Problematisch ist bei der internen Evaluation die zumeist nur gering ausgeprägte Methodenkompetenz und die fehlende Sicht von außen, die vor „Betriebsblindheit“ aufgrund fehlender Distanz bewahrt, und die z.B. bei der Lernerorientierten Qualitätstestierung durch externe Gutachter/innen gewährleistet wird (vgl. Abschnitt 3.1).

Den Nachteilen interner Evaluation stehen fast eins-zu-eins die Vorteile der externen Evaluation (in Abb. 3 diagonal) gegenüber: Die externe Evaluation wird meist von Personen mit hoher Methodenkompetenz durchgeführt und mit der Unabhängigkeit dieser Personen geht auch eine größere Glaubwürdigkeit einher, weshalb in diesem Zusammenhang auch von *überparteilicher Evaluation* die Rede ist (vgl. Gollwitzer & Jäger, 2007, S. 16). Allerdings haben in der Regel externe Evaluierende eine geringere Sachkenntnis über den Evaluationsgegenstand als interne Evaluierende. Darüber hinaus müssen sie vermehrt mit Abwehr in der Organisation rechnen, was sich für die Organisation auch in Umsetzungsproblemen ausdrücken kann. Scriven

<sup>13</sup> Dieses Beispiel zeigt, dass die Grenze zwischen interner und Selbstevaluation fließend ist, da auch bei einer internen Evaluation die am Gegenstand beteiligten Akteure Evaluationsaufgaben übernehmen können, z.B. der Professor an der Auswahl des Evaluationsinstruments, der Bewertung der Daten und der Verbreitung der Ergebnisse maßgeblich mitwirkt (vgl. auch DeGEval, 2004, S. 6).

differenziert in diesem Zusammenhang die *professionelle* und die *semi-professionelle Evaluation* und stellt indirekt einen Zusammenhang zu notwendigen Evaluationskompetenzen her: „Die professionelle Evaluation ist nach Scriven insbesondere durch ihre Unabhängigkeit vom evaluierten Gegenstand ausgezeichnet; eine Amateur-Evaluation wird dagegen von Personen durchgeführt, die in den Evaluationsgegenstand involviert sind“ (Kuper, 2005, S. 19).

Häufig besteht ein direkter Zusammenhang zwischen der Frage „Wer evaluiert?“ und den Evaluationszwecken. Verfolgt eine Evaluation primär die Entwicklung, Verbesserung und Optimierung eines Gegenstandes, werden sie häufig intern durchgeführt, während Evaluationen, die eine Legitimation bezwecken eher als externe Evaluation durchgeführt werden, „um möglichst hohe Objektivität und Glaubwürdigkeit zu erzielen“ (Stockmann, 2010c, S. 82).

### 1.6.3 Wie unabhängig ist die Evaluation?

Je nachdem, welche Rolle die Evaluierenden einnehmen, werden parteiliche von überparteilichen/objektiven Evaluationen unterschieden. Als *parteilich* bezeichnet man laut Gollwitzer und Jäger (2007, S. 12) eine Evaluation, wenn ein gewünschtes Ergebnis bereits zu Beginn der Evaluation wahrscheinlicher ist als ein anderes. Im Extremfall liegt bei einer parteilichen Evaluation das Evaluationsergebnis vor Durchführung der Evaluation fest, wobei in diesem Fall noch schwerlich von einer Evaluation gesprochen werden kann. Da die oben beschriebenen internen Evaluationen von Personen durchgeführt werden, die am Evaluationsgegenstand direkt beteiligt sind und diese meist ein Interesse daran haben, dass ihr Gegenstand positiv bewertet wird, unterliegen diese Evaluationen der Gefahr parteilich zu sein (vgl. ebd., S. 16). *Überparteiliche/objektive Evaluationen* liegen hingegen vor, wenn kein Ergebnis von vornherein wahrscheinlicher ist als ein anderes.

### 1.6.4 Wer erfährt von den Ergebnissen?

Eine Evaluation wird als *öffentlich* bezeichnet, wenn die Evaluationsergebnisse veröffentlicht werden, z.B. der Evaluationsbericht im Internet zum Download bereit steht. Die Veröffentlichung von Ergebnissen kann die Transparenz, Nachvollziehbarkeit und Glaubwürdigkeit steigern und wird beispielsweise bei politikrelevanten Evaluationen häufig gefordert (vgl. Stockmann & Meyer, 2010). Eine Evaluation ist *vertraulich*, wenn die Ergebnisse nur den Evaluierenden, den Auftraggebern und gegebenenfalls weiteren Beteiligten und Betroffenen der Evaluation bekannt sind,



wenn also die Ergebnisse nicht veröffentlicht werden. Der Großteil der Evaluationen ist vertraulich, weil die Ergebnisse nur für einen geringen Kreis an Personen überhaupt Relevanz besitzen, oder weil die Veröffentlichung von Evaluationsergebnissen die Marktposition schwächen könnte, etwa wenn die Evaluation sehr schlecht bewertete Kurse in einer Bildungsorganisation zu Tage fördert. Für die Differenzierung öffentlich versus vertraulich werden in der Evaluationsliteratur manchmal auch die Begriffe *offen* und *geschlossen* verwendet.

### 1.6.5 Welche Methoden und welche Designs kommen zum Einsatz?

Je nachdem, ob qualitative Methoden, wie nicht-standardisierte Interviews und qualitative Beobachtungen, oder quantitative Methoden, wie etwa standardisierte Fragebögen, eingesetzt werden, unterscheidet man *qualitative* und *quantitative Evaluationen*. Allerdings ist zu berücksichtigen, dass sich qualitative und quantitative Methoden nicht immer ganz trennscharf voneinander abgrenzen lassen und dementsprechend auch die Unterscheidung in qualitative und quantitative Evaluation nicht immer leicht zu treffen ist. Da in den letzten Jahren zunehmend die Mixed-Methods-Ansätze zur Kombination von qualitativen und quantitativen Daten (vgl. Kuckartz, 2009a) in der empirischen Sozialforschung an Bedeutung gewinnen und diese Kombination in der Evaluation häufig praktiziert und gefordert wird<sup>14</sup>, wird zunehmend häufiger von *Mixed-Methods-Evaluationen* gesprochen.

Bezüglich des verwendeten Untersuchungsdesigns ist zwischen vergleichender und nicht-vergleichender Evaluation zu unterscheiden, die Wulf bereits 1972 wie folgt vornimmt: „Die Evaluation mit Hilfe von Kontrollgruppen wird als *vergleichende Evaluation* bezeichnet. Bei Verzicht auf Kontrollgruppen spricht man von *nicht-vergleichender Evaluation*“ (Wulf, 1972, S. 402). Gollwitzer und Jäger (2007, S. 16) fassen diese Differenzierung etwas weiter und bezeichnen bereits den Vergleich verschiedener Maßnahmen, etwa bezüglich ihrer Konzeption, ihrer Umsetzung oder ihrer Wirkung, als vergleichende Evaluation.

Alle im Abschnitt 1.6 vorgestellten Unterscheidungsmerkmale sind in der Abb. 4 überblicksartig dargestellt.

---

<sup>14</sup> Beispielsweise wurde bei der 13. Jahrestagung der DeGEval im September 2010 ein Pre-Conference-Workshop zum Thema „Datenerhebung in der Evaluation: Notwendigkeit und Umsetzungsmöglichkeiten eines Methoden-Mix bei Evaluationen“ angeboten. Vgl. auch Kuckartz und Busch (2012) zu Mixed Methods in der Evaluation.

**Abb. 4: Unterscheidungsmerkmale von Evaluationen (eigene Darstellung)**

Wann wird evaluiert?	ex-ante, on-going, ex-post preformativ, formativ, summativ prognostisch, progressiv
Wer evaluiert?	Selbstevaluation, Fremdevaluation intern, extern semi-professionell, professionell
Wie unabhängig ist die Evaluation?	parteilich, überparteilich, objektiv
Wer erfährt von den Ergebnissen?	öffentlich, vertraut offen, geschlossen
Welche Methoden und welche Designs kommen zum Einsatz?	qualitativ, quantitativ, mixed-methods vergleichend, nicht-vergleichend

## 1.7 Kompetenzanforderungen an Evaluierende

Ausgehend von den USA und Kanada, in denen die Professionalisierung der Evaluationscommunity bereits weit voran geschritten ist, hat das Thema „Evaluationskompetenzen“ in den letzten Jahren auch im deutschsprachigen Raum und Europa stark an Bedeutung gewonnen. Im Zentrum der Überlegungen steht dabei zumeist die Frage, welche Kompetenzen Evaluierende benötigen, um Evaluationen erfolgreich durchzuführen („Evaluation Capacity“) und wie sie diese Kompetenzen erlangen („Evaluation Capacity Building“). Prominente Akteure des US-amerikanischen Evaluationsdiskurses haben das Thema in der Vergangenheit immer wieder in Einzelaktivität aufgegriffen. Mitte der 1990er Jahre listet beispielsweise Scriven (1996, S. 160) zehn „Necessary Competencies for Evaluators“ auf, die Evaluierende gut verstehen und anwenden können sollen. Hierzu zählen etwa Kenntnisse und Fertigkeiten in qualitativen und quantitativen Methoden, in der Bedarfsbewertung, in der Kostenanalyse und der evaluationsspezifischen Berichterstattung. Neben derartigen Einzelaktivitäten nehmen sich seit Mitte der 1990er Jahre insbesondere Evaluationsgesellschaften des Themas an und stellen die Diskussion auf eine breite Basis. So veröffentlicht die us-amerikanische Evaluationsgesellschaft 1995 die „Guiding Principles for Evaluators“ (American Evaluation Association, 1995) und vier Jahre später folgt die „Essential Skills Series in Evaluation“ der kanadischen Evaluationsgesellschaft (Canadian Evaluation Society, 1999). In den letzten zehn Jahren entstanden zudem in den USA, in Kanada, in Europa und in Deutschland unterschiedliche Kompetenzraster und Anforderungsprofile für Evaluierende, von denen eine Auswahl weiter unten in den Abschnitten 1.7.1 sowie 1.7.2 vorgestellt wird.

Im deutschsprachigen Raum beschäftigen sich seit der Jahrtausendwende verschiedene Publikationen mit dem Thema Evaluationskompetenzen. Im Jahr 2002 veröffentlicht Uwe Schmidt in der Zeitschrift für Evaluation einen Aufsatz über Methodenkompetenz in der Evaluation und Tasso Brandt legt eine Diplomarbeit mit dem Titel „Qualifikationsanforderungen für Evaluatoren – Überlegungen zur Entwicklung eines Ergänzungsstudiums Evaluation“ vor. 2004 erscheinen in erster und 2008 in zweiter, unveränderter Auflage die „Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatoreninnen und Evaluatoren“ des Arbeitskreises Aus- und Weiterbildung in der Evaluation (Caspari, Hennen, Scheffler, Schmidt & Schwab, 2008). Und im Jahr 2011 veröffentlicht die DeGEval ein Positionspapier zu Methoden der Evaluation, das einen kurzen Abschnitt zur Frage enthält, über welche Methodenkompetenz Evaluierende verfügen sollten:

„Evaluierende müssen über ein breites Methodenspektrum verfügen. Zwar kommt in keiner Evaluation das gesamte Methodenspektrum zum Einsatz. Doch die Evaluierenden müssen über genügend methodische Kenntnisse verfügen, um Potenziale und Grenzen der eingesetzten Verfahren und Zugänge einschätzen zu können sowie deren Anwendung entsprechend begründen zu können; was die Kenntnis alternativer Verfahren voraussetzt.“ (DeGEval, 2011, S. 2)

Auf paneuropäischer Ebene sind erst seit wenigen Jahren Aktivitäten zum Thema Evaluationskompetenzen zu beobachten und im Jahr 2009 hat die europäische Evaluationsgesellschaft EES die Ergebnisse einer Umfrage unter ihren Mitgliedern zu einem Rahmenwerk für Evaluationskompetenzen veröffentlicht (vgl. European Evaluation Society, 2009).

Der Nutzen, sich mit Evaluationskompetenzen auseinanderzusetzen, liegt im Wesentlichen darin, die Qualität von Evaluationen zu sichern und zu verbessern. So verwundert es nicht, dass im dritten Nützlichkeitsstandard der DeGEval ein enger Zusammenhang von Glaubwürdigkeit und Kompetenz der Evaluierenden zur Nützlichkeit von Evaluationen hergestellt wird: „Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird“ (DeGEval, 2008, S. 24). Perrin (2005, S. 171-176) hat die Diskussion um den Nutzen von der Beschäftigung mit Evaluationskompetenzen analysiert und beschreibt im Wesentlichen fünf Verwendungsbereiche: (1) *Ausbildung und Training*: Definierte Kompetenzen helfen unter anderem bei der Erstellung von Curricula. (2) *Selbstreflektion*: Anfänger/innen und Fortgeschrittene können ihre aktuell vorhan-

denen Kompetenzen einschätzen und Erweiterungs- sowie Auffrischungsbedarf feststellen. (3) *Professionalisierung*: Insgesamt kann die Beschäftigung mit Evaluationskompetenzen die Professionalisierung im Bereich Evaluation vorantreiben. (4) *Bewusstsein für die Notwendigkeit von Evaluierenden mit entsprechenden Evaluationskompetenzen* wird erhöht. (5) *Orientierung für Auftraggebende*: Evaluierende können entsprechend ihrer Kompetenzen ausgewählt werden.

Sich mit Evaluationskompetenzen zu beschäftigen, liegt aus einem weiteren Grund auf der Hand: Das Vorhandensein von Evaluationskompetenzen übt einen direkten Einfluss auf den Evaluationsprozess aus, denn nur, wer beispielsweise qualitative und quantitative Methoden sowie deren planvolle Kombination beherrscht, kann auch eine Mixed-Methods-Evaluation sinnvoll durchführen. Dem Ideal, dass gegenstandsangemessene Methoden bei der Evaluation zum Einsatz kommen (vgl. Stefer, in Vorbereitung), wie etwa der DeGEval-Evaluationsstandard D1 „Angemessene Verfahren“ fordert, stehen also die tatsächlich vorhandenen Kompetenzen der Evaluierenden gegenüber. Es entsteht gewissermaßen ein Kompetenzparadoxon: Eigentlich sollten nur die Rahmenbedingungen, Fragestellungen etc. ausschlaggebend für die angewendete Evaluationsmethode sein und nicht die Methodenkenntnisse und -präferenzen der Evaluierenden, doch sollten die Evaluierenden andererseits keine Methoden anwenden, die sie nicht beherrschen.

Nielsen, Lemire und Skov (2011) merken nach Durchsicht der aktuellen Literatur an, dass der Begriff Evaluationskompetenz auf drei Ebenen Anwendung findet: Auf der Mikroebene bezieht der Begriff sich auf das Individuum der Evaluatorin bzw. des Evaluators, auf der Mesoebene spricht er die Kompetenzen einer Organisation an und auf Makroebene die gesellschaftliche Ebene an. Die vorliegende Arbeit konzentriert sich auf die Mikroebene, wenngleich auch die Mesoebene zum Teil Berücksichtigung findet. Im Anschluss an die allgemeine Evaluationsdiskussion sowie die allgemeine Kompetenzdebatte<sup>15</sup> wird in dieser Arbeit unter Evaluationskompetenzen das Wissen, die Fertigkeiten und die Haltungen (knowledge, skills and dispositions) verstanden, welche Evaluierende benötigen, um Evaluationen durchzuführen (vgl.

---

15 Eine sehr lesenswerte, knappe Übersicht über den Kompetenzbegriff im Kontext der Bildung findet sich bei Zürcher (2010). Gnahs (2007) legt eine Monographie über den Erwerb und die Erfassung von Kompetenzen vor. Seine Kompetenzdefinition ist angelehnt an die der OECD (2003), die unter anderem auch den PISA-Studien zugrunde liegt, und lautet wie folgt: „Eine Kompetenz ist die Fähigkeit zur erfolgreichen Bewältigung komplexer Anforderungen in spezifischen Situationen. Kompetentes Handeln schließt den Einsatz von Wissen, von kognitiven und praktischen Fähigkeiten genauso ein wie soziale und Verhaltenskomponenten (Haltungen, Gefühle, Werte und Motivationen). Eine Kompetenz ist also zum Beispiel nicht reduzierbar auf ihre kognitive Dimension, sie beinhaltet mehr als das“ (Gnahs, 2007, S. 21-22).

Stevahn, King, Ghore & Minnema, 2005, S. 48). Der Arbeitskreis „Aus- und Weiterbildung in der Evaluation“ in der DeGEval hat 2011 ein Empfehlungspapier zur Didaktik der Evaluation vorgelegt, in dem diese drei Kompetenzbereiche weiter ausgeführt werden:

„*Wissen* umfasst inhaltliche Kenntnisse zum Themenbereich Evaluation (z.B. verschiedene Evaluationsansätze oder Stärken und Schwächen verschiedener Methoden). *Fertigkeiten* umfassen methodische Kenntnisse und praktische Fähigkeiten zur Planung, Umsetzung und Steuerung von Verfahren und Methoden (z.B. Stakeholderanalyse, Sampling oder Anregung der Evaluationsnutzung). *Haltungen* umfassen relevante Einstellungen und motivationale Tendenzen (z.B. Offenheit, Nutzungsorientierung oder Reflexionsbereitschaft).“ (Gutknecht-Gmeiner et al., 2011, S. 7, Aufzählungspunkte entfernt)

Wenn in den folgenden Abschnitten Kompetenzraster vorgestellt werden, sollte neben dieser Kompetenzdefinition auch berücksichtigt werden, an welche Art von Evaluierenden Kompetenzanforderungen gestellt werden. Denn in Abhängigkeit von der konkreten Evaluationstätigkeit und ihrem Kontext, insbesondere der Rolle, die die Evaluierenden übernehmen, ergeben sich höhere oder niedrigere Ansprüche. Dementsprechend unterscheiden die „Core Competencies for Evaluators of the UN System“ (UNEG, 2008) vier unterschiedliche Anspruchsniveaus vom Junior Officer (Level I) bis zum Senior Officer (Level IV). Ghore, King, Stevahn und Minnema (2006) schlagen zur Selbsteinschätzung von Evaluationskompetenzen die drei Stufen „Entry/Novice“, „Proficient/Skilled“ und „Mastery/Expert“ mit jeweils zwei Ausprägungen vor. Um die Bandbreite von Evaluierenden zumindest ansatzweise zu veranschaulichen, benennen Rädiker und Niklesz (2012) drei Dimensionen:

1. Evaluation als Teil- vs. Hauptaufgabe (Umfang),
2. Einfache vs. anspruchsvolle Evaluation (Komplexität, Qualitätsanspruch),
3. Interne vs. externe Evaluation (Rolle der Evaluierenden, Nähe zum Gegenstand).

Personen, die anspruchsvolle externe Evaluationen als Hauptaufgabe übernehmen, stehen deutlich höheren Anforderungen gegenüber als Personen, die einfache, interne Evaluationen als Teilaufgabe übernehmen.

### 1.7.1 Internationale Kompetenzraster

Wie eingangs berichtet, wurden in den letzten Jahren mehrere Kompetenzraster für Evaluierende entwickelt. Mit die größte Beachtung haben die „Essential Competencies for Program Evaluators (ECPE)“ nach Stevahn et al. (2005) gefunden, die für viele Weiterentwicklungen Pate standen. Die ECPE blicken auf eine längere Entwicklungsgeschichte zurück. Bereits Ende der 1990er Jahre hat die Gruppe um Jean A. King, Laurie Stevahn, Gail Ghore und Jane Minnema an der University of Minnesota ein erstes Set an Evaluationskompetenzen erarbeitet und auf einer Jahrestagung der American Evaluation Association (AEA) vorgestellt (vgl. King, Minnema, Ghore & Stevahn, 1998). In den Folgejahren wurde das Set unter anderem in Folge empirischer Untersuchungen fortgeschrieben (vgl. King, Stevahn, Ghore & Minnema, 2001, S. 231-244).

Die Essential Competencies gliedern sich in sechs Kompetenzkategorien auf: (1) Professionelle Praxis, (2) Systematische Untersuchung, (3) Situationsanalyse, (4) Projektmanagement, (5) Reflexive Praxis sowie (6) Interpersonelle Kompetenz. Jede Kategorie enthält unterschiedlich viele Teilkompetenzen, wobei exemplarisch jeweils die ersten drei Teilkompetenzen in Abb. 5 wiedergegeben sind. Die Teilkompetenzen sind jeweils als Indikatoren formuliert, die (mehr oder weniger) beobachtbares Verhalten beschreiben („nimmt teil“, „wendet an“, „setzt ein“, „versteht“ etc.).

**Abb. 5: Auszug aus den Essential Competencies for Program Evaluators (ECPE) (Quelle: Stevahn et al., 2005, S. 49-51; Ghore et al., 2006, S. 110; Übersetzung nach Beywl, 2006, S. 333-335 sowie eigene Übersetzung)**

<p><b>1. Professionelle Praxis:</b> Kompetenzen zur Anwendung von Normen und Werten, wie Evaluationsstandards und Ethikrichtlinien, die die Basis für eine professionelle Evaluationspraxis sind.</p> <p>1.1 Wendet professionelle Evaluationsstandards an.</p> <p>1.2 Handelt ethisch und strebt nach Integrität und Aufrichtigkeit bei der Durchführung von Evaluationen.</p> <p>1.3 Macht potenziellen Auftraggebern eigene Evaluationsansätze und -fertigkeiten deutlich.</p> <p>[..., insgesamt 6 Teilkompetenzen]</p>
<p><b>2. Systematische Untersuchung:</b> Technische bzw. methodische Kompetenzen, wie z.B. Evaluationsdesign, Datenanalyse und -interpretation sowie die Darstellung der Ergebnisse.</p> <p>2.1 Versteht die Wissensbasis von Evaluation (Begriffe, Konzepte, Theorien, Annahmen).</p> <p>2.2 Verfügt über Kenntnisse zu quantitativen Methoden.</p> <p>2.3 Verfügt über Kenntnisse zu qualitativen Methoden.</p> <p>[..., insgesamt 20 Teilkompetenzen]</p>
<p><b>3. Situationsanalyse:</b> Kompetenzen zur Analyse und Beachtung des (politischen) Kontexts einer Evaluation, einschließlich Evaluierbarkeit, Konfliktpotenzial und Aspekten des Nutzens.</p> <p>3.1 Beschreibt das Programm.</p> <p>3.2 Bestimmt die Evaluierbarkeit des Evaluationsgegenstands.</p> <p>3.3 Identifiziert die Interessen von relevanten Beteiligten und Betroffenen.</p> <p>[..., insgesamt 12 Teilkompetenzen]</p>
<p><b>4. Projektmanagement:</b> Kompetenzen im Hinblick auf eine projektförmige Gestaltung des Evaluationsprozesses von der Vertragsverhandlung bis hin zur zeitgerechten Durchführung.</p> <p>4.1 Antwortet auf Anforderungen zur Abgabe von Angeboten</p> <p>4.2 Führt mit Auftraggebern Verhandlungen, bevor die Evaluation beginnt.</p> <p>4.3 Setzt formelle Vereinbarungen auf.</p> <p>[..., insgesamt 12 Teilkompetenzen]</p>
<p><b>5. Reflexive Praxis:</b> Kompetenzen zur Wahrnehmung und Förderung der eigenen Professionalität.</p> <p>5.1 Nimmt sich bewusst als Evaluatorin/Evaluator wahr (Wissen, Fertigkeiten, Haltungen).</p> <p>5.2 Reflektiert die eigene Evaluationspraxis (Kompetenzen und Bereiche für Weiterentwicklung).</p> <p>5.3 Nimmt an beruflicher Weiterbildung teil.</p> <p>[..., insgesamt 5 Teilkompetenzen]</p>
<p><b>6. Interpersonelle Kompetenz:</b> Kompetenzen zur professionellen Gestaltung zwischenmenschlicher Interaktion, wie z.B. schriftliche/mündliche Kommunikation und Verhandlungstechniken.</p> <p>6.1 Setzt schriftliche Kommunikationsfertigkeiten ein.</p> <p>6.2 Setzt verbale/zuhörende Kommunikationsfertigkeiten ein.</p> <p>6.3 Setzt Verhandlungsfertigkeiten ein.</p> <p>[..., insgesamt 6 Teilkompetenzen]</p>

Das jüngste Kompetenzraster hat die kanadische Evaluationsgesellschaft „CES“ unter dem Namen „Competencies for Canadian Evaluation Practice (CCEP)“ vorgelegt. Es wurde in den Jahren 2008 und 2009 entwickelt und integriert neben anderen Quellen insbesondere die oben vorgestellten Essential Competencies for Program Evaluators (ECPE). Das kanadische Kompetenzraster ist den ECPE sehr ähnlich, umfasst allerdings nur fünf Kompetenzbereiche, weil die Kategorien „Reflective Practice“ und „Professional Practice“ zusammengelegt wurden. Auch die Teilkompetenzen wurden von 61 auf 49 durch Bündelung reduziert. Die entscheidende Weiterentwicklung des Rasters besteht in der umfassenden Konkretisierung der Teilkompetenzen. Wie die auszugsweise Darstellung in Abb. 6 verdeutlicht, wurde jede einzelne Teilkompetenz detailliert dahingehend ausgeführt, „what backgrounds, knowledge, skill and disposition is demonstrated by the evaluator in conducting these actions“ (Canadian Evaluation Society, 2010, S. 5). Die detaillierte Konkretisierung, die zu insgesamt 206 Einzelbeschreibungen geführt hat, ist unter anderem darauf zurückzuführen, dass in Kanada Evaluierende im Rahmen des „CES Professional Designations Programs“ zertifiziert werden, wobei der Abgleich mit den Einzelkompetenzen einen wichtigen Schritt im Zertifizierungsverfahren darstellt.<sup>16</sup>

**Abb. 6:** Auszug aus den Competencies for Canadian Evaluation Practice (CCEP) (Quelle: Canadian Evaluation Society, 2010, S. 5-9)

1.0 Reflective Practice	
...	
1.7 Pursues professional networks and self development to enhance evaluation practice	Be a member of an Evaluation professional organization Be involved in the work of an Evaluation Society Attend learning events (conference, courses, etc.) related to evaluation practice Undertake experiential and on the job learning activities
...	
2.0 Technical Practice	
...	
2.10 Collects Data	Collect information with a view to accuracy, neutrality, fairness, controlling bias and facilitating participation of respondents Address quality control issues, for example, conducts observer reliability checks, or takes steps regarding missing or incomplete program records Ensure adequate adherence to established data collection protocols (e.g., audit interviews to ensure that interviewers have been adequately trained in the required interview method) Determine extent of adherence to data entry protocols (i.e., draw a sample of entries and check against protocols and raw data) Negotiate access to data collection sites/populations Supervises data collection/ fieldwork
...	

16 Informationen zum CES Professional Designations Program finden sich unter:  
[http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&\\_lang=en](http://www.evaluationcanada.ca/site.cgi?s=5&ss=6&_lang=en) [10.12.2010].



Neben den USA und Kanada steht auch in Europa das Thema Evaluationskompetenzen auf der Agenda der Evaluationsgesellschaften. Allerdings ist die Entwicklung im Vergleich noch nicht sehr weit fortgeschritten. So hat das Netzwerk der europäischen Evaluationsgesellschaften auf der Konferenz der European Evaluation Society (EES) 2008 in Lissabon eine Initiative gegründet, die sich dem Thema annehmen soll (vgl. Picciotto, 2009, S. 1). Die Initiative hat eine Mitgliederbefragung durchgeführt und einen ersten Entwurf für eine Kompetenztaxonomie erarbeitet, die aus drei Hauptbereichen besteht: (1) Evaluation knowledge, (2) Professional practice, (3) Dispositions and attitudes (vgl. European Evaluation Society, 2009). Die drei Hauptbereiche greifen also keine für Evaluationen üblichen Inhaltsbereiche auf, sondern orientieren sich an üblichen Kompetenzdefinitionen, die Wissen, Fähigkeiten und Haltungen unterscheiden (siehe oben), was unter anderem sicherlich darauf zurückzuführen ist, dass es sich um ein „Rahmenwerk“ handeln soll. Insgesamt enthält das Rahmenwerk 30 Teilkompetenzen, wobei in Abb. 7 ein kurzer Auszug aus dem ersten Kompetenzbereich dargestellt ist.

**Abb. 7: Auszug aus dem Rahmenwerk der European Evaluation Society**  
(Quelle: European Evaluation Society, 2009, S. 21)

1. Evaluation knowledge
<b>1.1 Appreciates the role played by evaluation in diverse contexts</b>
1.11 Evinces familiarity with evaluation theory and approaches
1.12 Shows awareness of evaluation history and trends
1.13 Understands the linkages between evaluation and the social sciences
1.14 Grasps program theory and its evaluation implications
1.15 Relates evaluation to governance, policy and management environments in public, private and voluntary sectors
...

### 1.7.2 Das Kompetenzraster der DeGEval

Im deutschsprachigen Raum hat der DeGEval-Arbeitskreis „Aus- und Weiterbildung in der Evaluation“ im Jahr 2001 mit der Erarbeitung eines Anforderungsprofils an Evaluierende begonnen und nach Rückkoppelung mit den Mitgliedern der DeGEval im Jahr 2004 erstmalig die „Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren“ vorgelegt (Caspari, Hennen, Scheffler, Schmidt & Schwab, 2004). Im Jahr 2008 erschien eine zweite unveränderte Auflage der Empfehlungen. Wie der Titel der Publikation verrät, ist das Anforderungsprofil nicht wie in Kanada mit dem Blick auf Zertifizierung entwickelt worden, sondern insbesondere vor dem Hintergrund, eine professionelle Ausbildung in Evaluation zu befördern. Das Kompetenzraster unterscheidet vier

Kompetenzfelder, welchen jeweils mehrere Evaluationsstandards zugeordnet werden (vgl. ebd., S. 11):

1. Theorie und Geschichte der Evaluation (zugeordnete Standards: u.a. G3, G8, G9)
2. Methodenkompetenz (u.a. D1, D3, G4)
3. Organisations- und Feldkenntnisse (u.a. F1, F2, G1)
4. Sozial- und Selbstkompetenzen (u.a. D2, D3, F3)

Die vier Kompetenzfelder werden von einem querliegenden fünften Kompetenzfeld „Praxis der Evaluation“ begleitet, das Vorgaben zu dem Umfang von Praktika, der Qualität der Auszubildenden sowie der einzelnen Phasen von Evaluationen, in welchen die Lernenden Praxiserfahrungen sammeln sollen, enthält.

Alle fünf Kompetenzfelder sind in jeweils drei bis fünf Dimensionen aufgeteilt, die mit Beispielen illustriert und konkretisiert werden. In Abb. 8 sind die ersten vier Kompetenzfelder, ihre Dimensionen sowie jeweilige Konkretisierungen dargestellt.

**Abb. 8: Kompetenzfelder und Dimensionen gemäß der DeGEval-Empfehlungen in Stichworten**  
(Quelle: Caspari et al., 2008, S. 30-32)

1. EVALUATIONSTHEORIE UND -GESCHICHTE	
a) Grundlagen	<ul style="list-style-type: none"> <li>• Definitionen Evaluation und Evaluationsforschung</li> <li>• Kennzeichen einer Evaluation</li> <li>• Funktionen von Evaluationen</li> <li>• Dimensionen von Evaluationen</li> </ul>
b) Evaluationsgeschichte	<ul style="list-style-type: none"> <li>• Entwicklungstrends</li> <li>• Nationale Evaluationskulturen</li> <li>• Einflüsse von Kontextfaktoren auf die Entwicklung von Evaluation</li> </ul>
c) Evaluationsansätze	<ul style="list-style-type: none"> <li>• Theoretische und methodologische Ansätze und Modelle</li> <li>• Positionierung und Aufgabenstellung sowie methodische Ausgestaltung</li> </ul>
d) Evaluationsstandards	<ul style="list-style-type: none"> <li>• Sicherung der Qualität von Evaluation</li> <li>• Kommunikationsinstrument</li> <li>• Konfliktmanagement und Steuerung von Evaluationen</li> </ul>
2. METHODENKOMPETENZEN	
a) Grundzüge empirischer Sozialforschung, Untersuchungsdesign	<ul style="list-style-type: none"> <li>• Wissenschaftstheoretische Grundlagen</li> <li>• Entwicklung und Operationalisierung von Fragestellungen</li> <li>• Planung empirischer Untersuchungen, Auswahl- und Messverfahren</li> </ul>
b) Datenerhebung	<ul style="list-style-type: none"> <li>• Grundlagen von Erhebungsformen</li> <li>• Entwicklung von Erhebungsinstrumenten</li> </ul>
c) Statistische Kenntnisse	<ul style="list-style-type: none"> <li>• Univariate Häufigkeitsverteilungen, Kreuztabellierung, Varianzanalyse</li> <li>• Verfahren zur Messung von Zusammenhängen, Signifikanztests</li> </ul>
d) Datenverarbeitung, -auswertung, -interpretation	<ul style="list-style-type: none"> <li>• Anwendungskenntnisse in relevanten Softwarepaketen zur quantitativen und qualitativen Datenanalyse</li> <li>• Kodierung und Rekodierung</li> <li>• Dateninterpretation und Reporting</li> </ul>

<b>e) Projektorganisation</b>	<ul style="list-style-type: none"> <li>• Zeitplanung, Durchführungsplanung und -kontrolle</li> <li>• Kostenplanung und -kontrolle</li> <li>• Einführung in Fragen der Kosten-Leistungs-Rechnung</li> </ul>
<b>3. ORGANISATIONS- UND FELDKENNTNISSE</b>	
<b>a) Organisationswissen</b>	<ul style="list-style-type: none"> <li>• Organisationsbegriff, Organisieren</li> <li>• Handlung vs. Struktur</li> <li>• Kommunikation/Interaktion</li> </ul>
<b>b) Rechts- und Verwaltungswissen</b>	<ul style="list-style-type: none"> <li>• Einführung Rechtswissenschaft</li> <li>• Verwaltungslehre</li> <li>• Geschäftsprozesse</li> </ul>
<b>c) Spezifische Feldkenntnisse</b>	<ul style="list-style-type: none"> <li>• Unterschiedliche Praxisfelder</li> <li>• Spezifisches Organisations- und Kommunikationswissen</li> <li>• Spezifisches Rechts- und Verwaltungswissen</li> </ul>
<b>4. SOZIAL- UND SELBSTKOMPETENZEN</b>	
<b>a) Soziale Kompetenz</b>	<ul style="list-style-type: none"> <li>• Kontaktaufbau und -gestaltung</li> <li>• Perspektivenübernahme und Empathie</li> <li>• Feedback und Konfliktfähigkeit</li> </ul>
<b>b) Kommunikative Kompetenz</b>	<ul style="list-style-type: none"> <li>• Kommunikationstheorie</li> <li>• Praxis I: Zuhören und Sprechen</li> <li>• Praxis II: Lesen und Schreiben</li> </ul>
<b>c) Kooperationskompetenz</b>	<ul style="list-style-type: none"> <li>• Präsentieren und Moderieren</li> <li>• Gesprächs- und Verhandlungsführung</li> <li>• Kooperation und Arbeit in Gruppen</li> </ul>
<b>d) Selbstmanagementkompetenzen</b>	<ul style="list-style-type: none"> <li>• Motivation und Arbeitsstil</li> <li>• Auftrags-, Erwartungs- und Rollenklärung</li> </ul>
<b>e) Lern- und Problemlösekompetenz</b>	<ul style="list-style-type: none"> <li>• Reflexion und Fokussierung</li> <li>• Problemlösungsstrategien</li> <li>• Lernformen und -stile</li> </ul>

Im Gegensatz zu den weiter oben vorgestellten Anforderungsprofilen aus den USA und aus Kanada enthält das Anforderungsprofil der DeGEval keine outcome-orientierten Kompetenzformulierungen in Form von konkret beobachtbarem Verhalten, sondern konzentriert sich auf input-orientierte Inhalte, die im Rahmen einer Aus- und Weiterbildung vermittelt werden sollen (vgl. Picciotto, 2009, S. 7-8; Beywl, 2006, S. 322-323). Diese unterschiedliche Herangehensweise ist auf unterschiedliche Funktionen zurückzuführen, welche die Kompetenzraster übernehmen sollen. Während das DeGEval-Raster dezidiert als Empfehlungspapier für die Aus- und Fortbildung in Evaluation konzipiert ist, dienen die nordamerikanischen Ansätze eher als ein „Rating-, Steuerungsinstrument für die Kompetenzanalyse und -entwicklung“ (Beywl, 2006, S. 326) und werden in Kanada sogar für die Zertifizierung von Evaluierenden eingesetzt.

### 1.7.3 Konkretisierungen von Anforderungen an Evaluierende im Bildungsbereich

Während die bisher aufgeführten Anforderungsprofile aufgrund ihres intendierten Anwendungsspektrums zwangsläufig allgemein bleiben, existieren vereinzelte und verstreute Ausführungen, in denen Anforderungen an Evaluierende speziell für den Bildungsbereich, also dem Themenfeld dieser Arbeit, formuliert sind. Diese Ausarbeitungen schauen nicht immer aus einer dezidierten Evaluationsperspektive, greifen diese aber mehr oder weniger ausführlich auf. So listet Zech (2008) zahlreiche Kompetenzen auf, die zu einem Anforderungsprofil an Lehrende in Weiterbildungsorganisationen gehören können und versteht unter der methodischen Kompetenz unter anderem „Kenntnisse verschiedener Methoden der Bildungsarbeit und deren Evaluation“ (S. 79). Für den Bereich der Jugendhilfe, in dem wie in der Weiterbildung häufig auch interne bzw. Selbstevaluationen durchgeführt werden, haben Beywl, Mecklenburg, Richard, Schneid und Wonik (2001, S. 29-30) konkrete Empfehlungen vorgelegt, die unter anderem auch Qualifikationsanforderungen an Evaluierende beinhalten. Demnach sollen Evaluierende fachlich qualifiziert und persönlich kompetent sein, und insbesondere über folgende Kompetenzen verfügen:

- „– evaluationsfachliche Kenntnisse insbesondere zur Formulierung von Fragestellungen und Evaluationszielen, zu quantitativen Datenerhebungsinstrumenten sowie deskriptiver Statistik und zu qualitativen Erhebungsverfahren und zu ihrer Auswertung,
- Moderationsfertigkeiten, Fähigkeiten im Projektmanagement sowie Präsentationstechniken,
- kommunikative Kompetenz, Konfliktlösungstechniken, Verhandlungsgeschick sowie
- hinreichende Feldkenntnisse aus den jeweiligen Arbeitsfeldern der Jugendhilfe und auch quer zu Arbeitsfeldern, oder sie sollten sich diese kurzfristig aneignen, insbesondere um jugendhilfefachlich angemessene Fragestellungen und Bewertungsstandards formulieren zu können.“ (ebd., S. 30)

Hense und Mandl (2009, S. 134-135) beschäftigen sich mit den notwendigen Kompetenzen für die *institutionelle Selbstevaluation in der Schule*. Unter Rückgriff auf eine empirische Studie von Hense (2006) kommen die beiden Autoren zu dem Schluss, dass für die erfolgreiche Durchführung institutioneller Selbstevaluation neben Kenntnissen und Fertigkeiten in der Evaluation insbesondere eine Vielzahl an

konkreten Haltungen erforderlich ist: Positive Einstellungen im Sinne einer hohen Nutzenerwartung der Selbstevaluation, Abwesenheit von Befürchtungen hinsichtlich der Selbstevaluation, professionelle Neugier und Offenheit, konstruktive Problem- und Konfliktkultur, Verbesserungsorientierung und die Bereitschaft, sich kritischen Ergebnissen und Problembereichen zu stellen.

In der europäischen Studie „Key competences for adult learning professionals“ im Rahmen der Diskussion um die Professionalisierung von Weiterbildner/innen (vgl. Buiskool, Broek, Lakerveld, Zarifis & Osborne, 2010) wird die Kompetenz zur Evaluation zum einen als eine die Kernkompetenzen erweiternde Kompetenz gesehen, die im Wesentlichen auf den Lernprozess bezogen ist. Zum anderen werden Monitoring und Evaluation auch zu den „Overall management activities“ gezählt, welche die Organisationsstruktur und Qualitätssicherung betreffen (ebd., S. 38). Abb. 9 zeigt die detaillierten Erläuterungen bezüglich der Kompetenz „Being an Evaluator of the Learning Process“.

**Abb. 9: Kompetenzanforderung an adult learning professionals bezüglich Evaluation**  
(Quelle: Buiskool et al., 2010, S. 71)

**Title:** Competence to continuously monitor and evaluate the adult learning process in order to improve it on an ongoing basis: being an evaluator of the learning process.

**Description:** The adult learning professional is competent in continuously monitoring the progress of the adult learners in the learning process and in evaluating the learning process itself, the learning strategy chosen, their own professional practice and the learning resources used. The professional is able to assess the learning outcomes and is able to formulate a well-balanced and grounded judgement of the progress of the adult learners and is able to create an atmosphere in which adult learners feel free to give their judgement or opinion on the learning process and the role the professional plays in this.

*Essential knowledge, skills and attitudes are:*

**Knowledge:** The adult learning professional

- has knowledge of different monitoring and evaluation techniques
- has knowledge of the ways that outcomes can be used to improve the learning process, learning strategies and his/her own practice

**Skills:** The adult learning professional

- is able to use the different techniques
- is able to listen carefully
- is able to interpret the outcomes of the monitoring or evaluation process

**Attitudes:** The adult learning professional

- is self-reflective
- is willing to invest in the further development
- is willing to improve the learning process and the strategy used in his/her own practice



## **2 Evaluation in der Weiterbildung**

Nachdem in Kapitel 1 die allgemeinen Evaluationsgrundlagen im Vordergrund standen, konkretisiert Kapitel 2 die bisherigen Ausführungen für den Kontext der Weiterbildung. Da in dieser Arbeit die Evaluation von Weiterbildungsprozessen empirisch untersucht wird, ist es notwendig, das zugrundeliegende Verständnis von Weiterbildung, Weiterbildungsprozessen und Evaluation auszuweisen. Dies geschieht in den Abschnitten 2.1 und 2.2. Zudem erscheint es sinnvoll, das Konzept der Evaluation gegenüber den artverwandten Konzepten „Bildungscontrolling“ und „Qualitätsmanagement“ abzugrenzen (Abschnitt 2.3). Die sich anschließenden Unterkapitel widmen sich zentralen Unterscheidungsmerkmalen von Evaluationen und verwendeten Evaluationsmethoden (Abschnitte 2.4 und 2.5). Als Vorbereitung der empirischen Untersuchung, die neben dem Status quo auch Herausforderungen in der Evaluationspraxis thematisiert, endet das Kapitel mit einer Darstellung von Evaluationsansätzen, die für die Weiterbildung Relevanz haben, und stellt abschließend zentrale Herausforderungen vor, die im Rahmen einer Evaluation auftreten können (Abschnitte 2.6 und 2.7).

### **2.1 Weiterbildung und Weiterbildungsprozesse: eine relevante Begriffsklärung**

Dieser Arbeit liegt eine Definition von Weiterbildung zugrunde, die sich wie das Berichtssystem Weiterbildung IX (vgl. Kuwan, Bilger, Gnahs & Seidel, 2006, S. 12) an der breiten Definition des Deutschen Bildungsrates von 1970 orientiert. Demnach wird Weiterbildung verstanden als:

„Fortsetzung oder Wiederaufnahme organisierten Lernens nach Abschluß einer unterschiedlich ausgedehnten ersten Ausbildungsphase ... Das Ende der ersten Bildungsphase und damit der Beginn möglicher Weiterbildung ist in der Regel durch den Eintritt in die volle Erwerbstätigkeit gekennzeichnet ... Das kurzfristige Anlernen oder Einarbeiten am Arbeitsplatz gehört nicht in den Rahmen der Weiterbildung.“ (Deutscher Bildungsrat, 1970, S. 197)

Konkret sollen hier unter Weiterbildung alle Aktivitäten gefasst werden, die der Vertiefung, Erweiterung oder Erneuerung von Wissen, Fähigkeiten und Fertigkeiten dienen (vgl. Nagel & Tiedtke, 2007, S. 3). Allerdings ist aufgrund des empirischen Fokus dieser Arbeit auf Weiterbildungsorganisationen diesbezüglich eine Einschränkung vorzunehmen: Die in jüngster Zeit verstärkt diskutierten und schwierigen

ger zu fassenden Lernformen des informellen Lernens und des Selbstlernens finden nur insofern Berücksichtigung, als sie unter Organisation und Betreuung durch eine Weiterbildungsorganisation stattfinden und nicht gänzlich selbstorganisiert sind.

Die Begriffe Erwachsenenbildung und Weiterbildung werden in dieser Arbeit wie auch in vielen Artikeln im „Handbuch Erwachsenenbildung/Weiterbildung“ von Tippelt und von Hippel (2010) synonym verwendet, wenngleich einige Autoren/innen wichtige Unterschiede markieren und vor allem aus einer historisch analytischen und kritisch reflektierenden Sichtweise „Erwachsenenbildung“ als etwas Umfassenderes und weniger Instrumentalisiertes betrachten als „Weiterbildung“ (vgl. z.B. Weinberg, 2000, S. 9-12).

Der sehr weite und schillernde Begriff „(Weiter-)Bildungsprozess“<sup>17</sup> ist etwas schwieriger zu fassen, da er in der Literatur wie selbstverständlich verwendet wird, ohne genauer bestimmt zu sein.<sup>18</sup> So wie *Bildung* üblicherweise eine umfassendere Bedeutung hat als *Lernen und Lehren*, so soll auch in dieser Arbeit der Begriff des Bildungsprozesses nicht gleichgesetzt werden mit einem Lehr-Lern-Prozess, sondern darüber hinaus gehen. Denn bei der Betrachtung von Lehr-Lern-Prozessen sind meist nur konkrete Lernsituationen und Lernarrangements im Fokus der Aufmerksamkeit. Bildungsprozesse umfassen jedoch auch die organisatorischen Abläufe, die das Lehren und Lernen rahmen und dem Lehr-Lern-Prozess meist vor- und nachgelagert sind. Bei der Betrachtung eines Bildungsprozesses stellt es einen relevanten Unterschied dar, ob man diesen aus der Perspektive eines (potenziellen) Teilnehmenden bzw. Lernenden oder aus der Perspektive einer Weiterbildungsorganisation betrachtet. Für die Lernenden beginnt ein Bildungsprozess mit der Bewusstwerdung eines Lernbedürfnisses und erstreckt sich über den ersten Kontakt mit einer Bildungsorganisation (oder sogar mehreren), der eigentlichen Bildungsveranstaltung bis hin zum Transfer und zur Adaption des Gelernten in den privaten und beruflichen Alltag. Für die Weiterbildungsorganisationen beginnt der Bildungsprozess mit der Bedarfserschließung, also der Analyse vorhandener Lernbedürfnisse und Marktchancen, über die Programmentwicklung, Lehrendenauswahl und Bewerbung der Angebote bis hin zur Realisierung der Veranstaltungen, deren Abschluss, even-

---

17 Im Folgenden wird häufiger nur von „Bildungsprozess“ die Rede sein, gemeint ist aber immer der „Weiterbildungsprozess“.

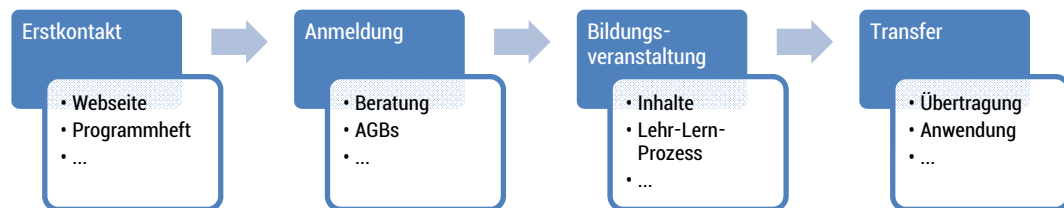
18 Am ehesten finden sich Abhandlungen über „Bildungsprozesse“ in der Fachliteratur zum Bildungscontrolling. Schöni (2009) berichtet beispielsweise, dass Seeber (2000, S. 35-36) ein Bildungsprozessmodell vorgelegt hat, „das Bildungsprozesse und ihre Teilschritte analog der Wertschöpfungskette von Produktionsbetrieben modelliert und den Fokus auf die Effektivität und Professionalität legt“ (Schöni, 2009, S. 44).



tuell der Unterstützung des Transfers sowie der abschließenden Auswertung der Erfolge. Dementsprechend nennt Schöni (2009, S. 38) für den Kernprozess der Bildung und Beratung vier Schritte: (1) Bedarfsabklärung bzw. Standortbestimmung (2) Planung und Konzeption von Programmen, Maßnahmen, Lernschritten, Beratungen (3) Realisierung und Umsetzung sowie (4) Programmevaluation und Erfolgskontrolle.<sup>19</sup>

In Abb. 10 sind die Schritte des Bildungsprozesses dargestellt, bei denen die Lernenden und die Weiterbildungsorganisation miteinander in persönlichen oder virtuellen Kontakt treten. Es handelt sich also um die Schritte, bei denen die Perspektiven von den Lernenden und den Organisationen zusammenfallen. Zu jedem Schritt sind einige Beispielaspekte genannt, welche diesen Prozessschritt kennzeichnen und im Fokus einer Evaluation stehen können. Im Gegensatz zu den ersten drei Schritten ist der vierte Schritt „Transfer“ insbesondere für Unternehmen der beruflichen Bildung bedeutsam, seltener für Organisationen der allgemeinen Weiterbildung und andere Bildungseinrichtungen.

**Abb. 10: Schritte des gemeinsamen Weiterbildungsprozesses von Lernenden und Organisation (eigene Darstellung)**



Für die vorliegende Arbeit erscheint es im Hinblick auf das in Kapitel 3 beschriebene LQW-Modell angemessen, zusätzlich zu den einzelnen Schritten eines Weiterbildungsprozesses vier unterschiedliche Dimensionen zu unterscheiden (vgl. auch Rädiker, 2012):

1. eine Dimension des Lernens,
2. eine des Lehrens,
3. eine der Lerninfrastruktur und
4. eine der Lernorganisation.

<sup>19</sup> In der LQW-Systematik (vgl. Kapitel 3) sind für den Qualitätsbereich zur Evaluation von Bildungsprozessen vorwiegend die Schritte bedeutsam, an denen die Lernenden direkt beteiligt sind und eine Interaktion zwischen Lernenden und Organisation stattfindet. Alle anderen Aspekte wie etwa die Bedarfserschließung oder die Programmerstellung werden bei LQW in anderen Qualitätsbereichen fokussiert.

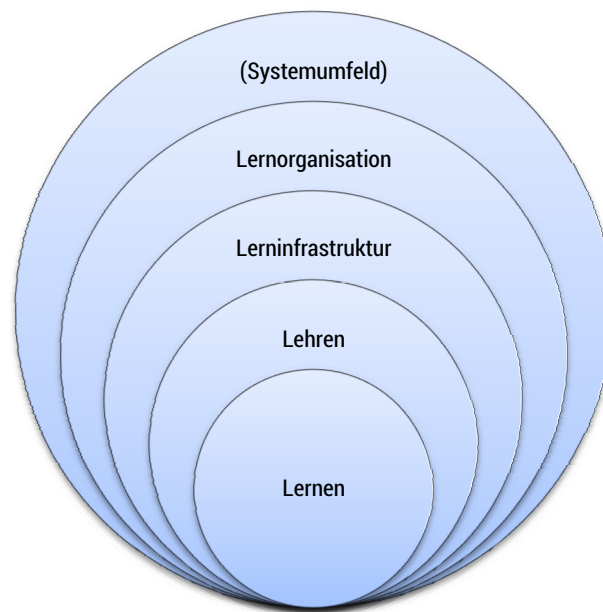
Mit *Lernen* ist das Aneignen von Wissen, Fertigkeiten und Haltungen eines Menschen gemeint und umfasst den Prozess des „Sich-Bildens“ bei den Lernenden. Das *Lehren* bezieht sich auf alle Aspekte, welche die Lehrpersonen und ihr Lehrhandeln betreffen, von ihrem fachlichen Wissen über ihre didaktisch-methodischen Umsetzungen bis hin zu ihrem Kommunikationsverhalten. Die Dimension der *Lerninfrastruktur* bezieht sich auf Räume, Mobiliar, aber auch auf die eingesetzten Lernmaterialien und Medien. Die vierte Dimension der *Lehrorganisation* rahmt die anderen Komponenten; sie umfasst unter anderem die Anmeldung und Beratung, die Information der Teilnehmenden im Vorfeld sowie die gesamte Kommunikation mit den Teilnehmenden außerhalb des direkten Lehr-Lern-Settings. Neben diesen vier Dimensionen ist es hilfreich zu berücksichtigen, dass Bildungsprozesse grundsätzlich in ein Systemumfeld eingebettet sind:

„Bei der Analyse und Steuerung von Bildungsprozessen muss das Bildungscontrolling den Arbeitsmarkttrends, der Bildungspolitik, den Kundenbedürfnissen, den Zielvorgaben von Auftraggebern und Aufsichtsinstanzen, den Erwartungen von Zielgruppen und gesellschaftlichen Anspruchsgruppen besondere Aufmerksamkeit widmen. Dies gilt grundsätzlich für die Personalentwicklung oder Bildungsabteilung in Unternehmen ebenso wie für selbstständige Bildungsanbieter. Bildungsanbieter öffentlichen Rechts sind bei der Gestaltung ihres Angebots zudem an einen staatlichen Bildungsauftrag gebunden.“ (Schöni, 2009, S. 37)

Die vorgestellten vier Dimensionen des Bildungsprozesses, ergänzt um die fünfte Dimension des Systemumfelds, sind in Abb. 11 dargestellt.<sup>20</sup> Je nachdem auf welche Dimension sich eine Evaluation richtet, liefert sie unterschiedliche Ergebnisse. Evaluationen auf der Dimension „Lernen“ fokussieren die Teilnehmenden und beurteilen etwa die Lernerfolge oder die Möglichkeit, Einfluss zu nehmen auf den Lernprozess. Evaluationen auf der Dimension „Lehren“ fokussieren die Lehrenden und beurteilen etwa die didaktische Steuerung oder den Medieneinsatz. Evaluationen auf der Dimension „Lerninfrastruktur“ fokussieren die Lernumgebung und die Lernhilfsmittel und beurteilen etwa die Räumlichkeiten und die Pausenverpflegung. Evaluationen auf der Dimension „Lernorganisation“ fokussieren die organisatorische Steuerung und beurteilen etwa das Anmeldeprocedere oder die Hinweise auf einen Veranstaltungsort. Evaluationen auf der Ebene des Systemumfelds fokussieren die Kontextbedingungen und beurteilen etwa die Bedingungen des Standorts oder die Erwartungen der Lokalpolitik.

---

<sup>20</sup> Weil das Systemumfeld als ergänzender Kontext betrachtet wird und im weiteren Verlauf der Arbeit nur bedingt berücksichtigt wird, steht es in Klammern.

**Abb. 11: Dimensionen des Weiterbildungsprozesses (eigene Darstellung)**

Wenn man die weiter oben in Abb. 10 vorgestellten Schritte eines Bildungsprozesses mit der Unterteilung in fünf Dimensionen aus Abb. 11 kombiniert, erhält man eine zweidimensionale Matrix, in der sich die jeweiligen Schritte weiter ausdifferenzieren lassen. Die einzelnen Dimensionen werden bei den einzelnen Schritten eine unterschiedlich ausgeprägte Bedeutung im Rahmen einer Evaluation haben und dementsprechend stark oder weniger stark in den Vordergrund treten. Während etwa beim Erstkontakt und der Anmeldung vorrangig Aspekte der Lernorganisation evaluiert werden wird, spielen bei der Bildungsveranstaltung besonders die Dimensionen Lernen und Lehren eine Rolle. Diese unterschiedlich ausgeprägten Bedeutungen werden in Abb. 12 durch Plus-Zeichen verdeutlicht: Je mehr Plus-Zeichen eine Zelle enthält, desto größere Bedeutung hat die Dimension für den jeweiligen Schritt.

**Abb. 12: Weiterbildungsprozess aus Sicht einer Organisation differenziert nach vier Dimensionen (eigene Darstellung)**

	<i>Bedarfs-erschließung*</i>	<i>Planung, Entwicklung*</i>	Erstkontakt	Anmeldung	Bildungs-veranstaltung	Transfer
Lernen	++	+++			++++	++++
Lehren	++	+++			++++	+
Lerninfrastruktur	+	+++	+	+	+++	+
Lernorganisation	+	+	+++	+++	++	+
(umgebendes System)	+++	++			+	+

Die Anzahl der Plus-Zeichen entspricht der Bedeutung der Dimension für den jeweiligen Schritt.

\* Die ersten beiden Schritte finden in der Regel ohne die Lernenden statt und sind deshalb *kursiv* gesetzt.

## 2.2 Evaluation von Weiterbildungsprozessen

Der kurze historische Abriss im Abschnitt 1.3 hat bereits gezeigt, dass Evaluationen stets eng verbunden waren mit dem Bildungsbereich, der zahlreiche Impulse für die Entwicklung der Evaluationsprofession beigesteuert hat. Hier sei nur noch einmal exemplarisch an die zahlreichen Bildungsbemühungen in Deutschland in den 1970er Jahren erinnert, die umfassend evaluiert wurden. Mit dem Aufkommen der Evaluation im Bildungswesen hat auch die Evaluation in der Weiterbildung Einzug gehalten. Tietgens (1971) beschreibt diese Entwicklung in der Einleitung des Abschnitts „Methoden > Arbeitskontrollen“ im Praxishandbuch für die Volkshochschule wie folgt:

„Solange die EB [Erwachsenenbildung] in der Improvisation und der Unverbindlichkeit besondere Vorzüge sah, stellte sich die Frage nach der Arbeitskontrolle nicht. Seitdem sie sich aber als Teil des Bildungssystems versteht und gezielt den Teilnehmerbedürfnissen gerecht werden will, muß sie auch versuchen, etwas über ihre Effizienz in Erfahrung zu bringen.“ (Blatt 73.000)

Im Laufe der Zeit hat der Begriff „Evaluation“ artverwandte Begriffe wie „Lernkontrolle“, „Arbeitskontrolle“ und „Erfolgskontrolle“ verdrängt. So stellt Wesseler (2010, S. 1037) fest, dass im Handbuch der Erwachsenenbildung von Pöggeler in den Jahren seines Erscheinens von 1974-1981 nur das Wort Kontrolle, nicht aber das Wort Evaluation im Register zu finden ist. Aus aktuellen Handreichungen und Nachschlagewerken zur Weiterbildung ist der Begriff Evaluation nicht mehr wegzudenken. So enthält das 2001 erschienene „Wörterbuch Erwachsenenpädagogik“ von Arnold, Nolda und Nuissl eine kurze Abhandlung über Evaluation in der Weiterbildung und das Einführungsbuch zur Erwachsenenbildung für Bachelor und Master von Faulstich und Zeuner widmet 2010 der Evaluation ein eigenes umfassendes Kapitel.

Es wäre vermessen zu behaupten, vor der Ausbreitung des Evaluationsbegriffs sei dem pädagogischen Personal die Vergewisserung über den Erfolg der eigenen Arbeit gleichgültig gewesen. Das Gegenteil ist sicherlich der Fall (vgl. Will et al., 1987b, S. 15; Reischmann, 2006, S. 86-88), doch zwingt in jüngster Zeit die zunehmende Diskussion um Qualität, in dessen Kontext auch das in dieser Arbeit behandelte LQW-Modell steht, zu einer systematischen Reflexion über die Evaluationspraxis. Darüber hinaus hat sich die Theorie und Praxis der Evaluation in Deutschland in den letzten 40 Jahren stark weiterentwickelt und diese Entwicklungen sind nicht spurlos an der Weiterbildungslandschaft vorübergezogen.

Beschäftigt man sich mit der Evaluation in der Weiterbildung, ist zu beachten, dass in der Diskussion häufig sehr unterschiedliche Ebene miteinander vermengt werden. Deshalb wurden in der folgenden Abb. 13 drei Ebenen voneinander abgegrenzt (vgl. Rädiker, 2009, S. 139).

**Abb. 13: Evaluation in der Weiterbildung auf verschiedenen Ebenen (eigene Darstellung)**

Evaluationsebene	Beispiel
Makro: Weiterbildungslandschaft	Evaluation zur Überprüfung der Wirksamkeit von Bildungsschecks
Meso: Weiterbildungsorganisation	Evaluation zur Verbesserung von Organisationsabläufen
Mikro: Weiterbildungsprozess	Evaluation zur Optimierung einer Bildungsveranstaltung

Auf oberster Ebene ist die Evaluation der Weiterbildungslandschaft als Ganzes, also des Gesamtsystems Weiterbildung angesiedelt. Diese Makroebene wird häufig von der Politik adressiert und in den Blick genommen. Auf der mittleren Mesoebene befinden sich Evaluationen, die sich auf einzelne Weiterbildungsorganisationen beziehen, etwa wenn eine Organisation eine Selbst- oder Fremdevaluation durchführt, deren Zweck meist in einer Mischung aus Selbstvergewisserung, Organisationsentwicklung und Legitimation liegt. Auf der untersten (Mikro-)Ebene liegt die Evaluation von Weiterbildungsprozessen, die im Mittelpunkt dieser Arbeit steht. Das LQW-Modell ist als Organisationsentwicklungsverfahren mit Anteilen der Selbst- und Fremdevaluation auf der zweiten Ebene zu verorten, während der im LQW-Modell enthaltene Qualitätsbereich „Evaluation von Bildungsprozessen“ überwiegend der untersten Ebene zuzuordnen ist, wobei er natürlich auch die Mesoebene der Organisationsprozesse direkt tangiert (vgl. hierzu die Ausführungen in Kapitel 3).

Wesseler (2010, S. 1037) unterscheidet nahezu analog zu den hier vorgestellten Ebenen die Evaluation von Programmen, Institutionen und dem übergeordneten „System“. Er benennt darüber hinaus auch noch die Evaluation von Fähigkeiten, Verhalten, Leistungen etc. einer Person. Gutknecht-Gmeiner (2009) betitelt einen Aufsatz mit „Evaluation (in) der Weiterbildung“ und bringt mit der Unterscheidung von „Evaluation *der* Weiterbildung“ und „Evaluation *in der* Weiterbildung“ die häufig anzutreffende Ebenenvermischung zum Ausdruck. Die Autorin unterscheidet ähnlich wie in Abb. 13 die Mikro-, Meso- und Makroebene, ergänzt diese jedoch noch um eine vierte, gewissermaßen querliegende Ebene, auf der „Modellvorhaben“ angesiedelt sind. Jede dieser Ebenen ist mit typischen Evaluationsgegenständen verknüpft, die in der zweiten Spalte von Abb. 14 aufgeführt sind.

**Abb. 14: Unterschiedliche Evaluationsgegenstände auf unterschiedlichen Evaluationsebenen (nach Gutknecht-Gmeiner, 2009, S. 13–6)**

<b>Ebene (nicht immer gänzlich trennscharf)</b>	<b>Mögliche Evaluationsgegenstände (exemplarisch)</b>
Modellvorhaben auf verschiedenen Ebenen	Entwicklung von Angeboten für bestimmte Zielgruppen/Regionen, Vernetzung und Austausch zwischen Institutionen etc.
Makroebene: Erwachsenenbildung/Weiterbildung als Ganzes oder Teilbereiche davon	Weiterbildungsangebot, Qualität, Beteiligung nach Zielgruppen, Ressourcen und Finanzierung, Wirkung von Weiterbildung (Bildung von Humankapital) etc.
Mesoebene: Institution	Aufbau- und Ablauforganisation (Strukturen und Prozesse), Programm(entwicklung); Kompetenz und Entwicklung der MitarbeiterInnen, Organisationsentwicklung und Innovation, Infrastruktur und finanzielle Ressourcen etc.
Mikroebene: Unterricht	Inhalte, Didaktik Interaktionen, Lehrende, Rahmenbedingungen, Ergebnisse, Transfer und Wirkungen

Will et al. (1987b, S. 17) unterscheiden ebenfalls verschiedene Ebenen der Handlungssteuerung durch Evaluation:<sup>21</sup>

1. Entwicklung einer Konzeption für das Gesamtangebot (z.B. übergeordnete Ziele, Häufigkeit einzelner Angebote)
2. Planung und Gestaltung einer in sich geschlossenen Bildungsveranstaltung
3. Vorbereitung und Realisierung eines Seminarbausteins
4. Durchführung eines konkreten didaktischen Schrittes

Die Unterscheidung fokussiert im ersten Punkt die Mesoebene der Organisation und differenziert die Mikroebene in drei Bereiche aus. In der Praxis ist weder die Aufgliederung in drei noch in vier Ebenen trennscharf, da die einzelnen Ebenen teilweise miteinander verschränkt sind. So kann die Gestaltung von Bildungsprozessen abhängig von Organisationsabläufen sein, während die Organisation ihrerseits durch das Wirkungsgefüge in der Weiterbildungslandschaft beeinflusst wird.

Was genau ist unter Evaluation in der Weiterbildung – bezogen auf die Mikro- und Mesoebene der Bildungsprozesse – zu verstehen? Zur Klärung dieser Frage werden im Folgenden die Definitionen von drei Autoren herangezogen. Gerl (1983, S. 19) definiert Evaluation in seiner Abhandlung über „Evaluation in Lernsituationen“ wie folgt: „Evaluation sollen alle jene Handlungen heißen, die dazu dienen, den Grad der Reflexivität von oder in Lernsituationen zu erhöhen.“ Dieses Verständnis führt er weiter aus:

---

<sup>21</sup> Die Unterscheidung bezieht sich zwar vornehmlich auf die berufliche Aus- und Weiterbildung, lässt sich jedoch aufgrund der allgemein gehaltenen Formulierungen problemlos auf andere Weiterbildungsbereiche übertragen.

„Evaluieren heißt jetzt, Klarheit darüber herzustellen,

- welches Ziel Kursleiter oder Teilnehmer mit ihrem gemeinsamen Handeln verfolgen,
- inwiefern Kursleiter oder Teilnehmer durch ihr Tun handlungsfähiger geworden sind,
- inwieweit die Formen der sozialen Interaktion in der Lernsituation als lernförderlich gelten können.

Dies sind Fragen, mit deren Hilfe das Lernen in Lernsituationen gemessen werden soll. Ziel einer jeden Evaluation ist, durch die mit ihrer Hilfe herbeigeführte Aufklärung die Lernqualität solcher Situationen zu verbessern.“ (ebd., S. 20)

Gerls Definition legt den Fokus auf die Mikrodidaktik, bezieht sich direkt auf den Lehr-Lern-Prozess und betont die vielfältigen Verbesserungsfunktionen von Evaluation, wobei im Mittelpunkt die reflexive Komponente der Evaluation steht. In der Definition selbst lässt Gerl offen, mit welchen Mitteln und Instrumenten die Optimierung vonstattengehen soll. Jedoch führt er im weiteren Verlauf seiner Ausführungen in Form zahlreicher Kriterien seinen Anspruch an Evaluationsmethoden aus und stellt zudem sowohl Evaluationsmethoden vor, die unterstützend in das Lerngeschehen eingreifen, als auch abschließend die Lernergebnisse bewerten. Sozialwissenschaftlichen Charakter weisen allerdings nur wenige der Methoden auf. Zum Abschluss seiner Überlegungen weist Gerl darauf hin, dass Evaluation als soziales Handeln zu verstehen ist und er führt resümierend aus: „Evaluation ist der Terminus für alle Bemühungen, die Situation in Lerngruppen allen Beteiligten transparent und in ihrer Entwicklung handhabbar zu machen“ (ebd., S. 88).

Die zweite hier präsentierte Definition stammt aus dem LQW-Modell, in dem die Evaluation von Bildungsprozessen als einer von mehreren Qualitätsbereichen eine wichtige Rolle einnimmt. Der LQW-Leitfaden für die Praxis enthält folgende für die Qualitätsentwicklung handlungsleitende Definition:

„Evaluation von Bildungsprozessen bedeutet, dass die durchgeführte Bildungsarbeit regelmäßig mit geeigneten Instrumenten geprüft und bewertet wird. Maßstabsbildend zur Bewertung sind der Lernerfolg, die Zufriedenheit der Teilnehmenden und ggf. der Auftraggeber sowie die Realisierung des eigenen institutionellen Anspruchs. Auch die Einschätzung der Lehrenden sollte Teil der Evaluation sein.“ (Zech, 2006, S. 68)

Während die Definition von Gerl tendenziell aus der Sicht von Kursleitenden formuliert ist, nimmt die LQW-Definition von Zech die Perspektive der Weiterbildungsorganisation als Ausgangspunkt. Auch diese Definition ist in Einzelbereichen sehr konkret und benennt beispielsweise maßstabsbildende Bewertungskriterien, wie den Lernerfolg und die Zufriedenheit der Teilnehmenden. Darüber hinaus führt Zech explizit die Lehrenden als Stakeholder des Evaluationsprozesses an, wobei der Formulierung „Einschätzung der Lehrenden“ im letzten Satz der Definition eine dreifache Bedeutung innewohnt: Erstens die Einschätzung der Lehrtätigkeit durch die Organisation, zweitens die Einschätzung der Lehrtätigkeit durch die Lehrenden selbst und drittens die Einschätzung von Evaluationsergebnissen durch die Lehrenden (vgl. Zech, 2008, S. 93). Die etwas vage Festlegung in der Definition auf „geeignete Instrumente“ zur Evaluation ist auf die Eigenheit des LQW-Modells zurückzuführen, das versucht auf die Vorschreibung von konkreten normativ-gesetzten Verfahren zu verzichten und stattdessen einen Rahmen aufspannt, den die Organisationen durch Reflexion und Begründung jeweils organisationsspezifisch füllen müssen (vgl. Kapitel 3).

Als dritte Definition sollen noch die Überlegungen von Reischmann (2006) herangezogen werden, der mit seinem Buch „Weiterbildungs-Evaluation“ eine der wenigen Monographien zum Thema vorgelegt hat. Seine Definition ist im Vergleich zur LQW-Definition etwas weiter gefasst:

„Evaluation meint

1. das methodische Erfassen,
2. das begründete Bewerten von Prozessen und Ergebnissen zum
3. besseren Verstehen und Gestalten einer Praxis-Maßnahme im Bildungsbereich durch Wirkungskontrolle, Steuerung und Reflexion.“ (S. 18)<sup>22</sup>

---

22 Will et al. (1987b, S. 14) benennen für den Kontext der beruflichen Bildung bereits einige Jahre zuvor ganz ähnliche Merkmale von Evaluationen wie Reischmann: (1) Evaluation ist ziel- und zweckorientiert. Sie hat vor allem das Ziel, Bildungsmaßnahmen zu verbessern, zu legitimieren oder über sie zu entscheiden, (2) Grundlage der Evaluation ist eine systematisch gewonnene Datenbasis über Voraussetzungen, Kontext, Prozesse und Wirkungen einer Maßnahme, (3) Evaluation beinhaltet eine bewertende Stellungnahme, das heißt die methodisch gewonnenen Daten und Ergebnisse werden auf dem Hintergrund von Wertmaßstäben unter Anwendung bestimmter Regeln bewertet, (4) Evaluation zielt in der Regel nicht primär auf die Bewertung des Verhaltens (z.B. Leistungen) einzelner Personen, sondern ist Bestandteil der Entwicklung, Realisierung und Kontrolle planvoller Bildungsarbeit. (Der vierte Punkt erscheint allerdings problematisch vor dem Hintergrund, dass Evaluationen häufig auch das Kursleiterverhalten adressieren.)



Im Vergleich zu den zuvor angeführten Definitionen kommt diese dritte Definition dem Verständnis der allgemeinen Debatte um Evaluation am nächsten: Es wird konkret das begründete Bewerten benannt, die Formulierung „das methodische Erfassen“ verweist auf eine systematische Datenerhebung bzw. Informationssammlung und zudem wird auch der Zweck von Evaluationen konkretisiert. Um seine Definition zu präzisieren, führt Reischmann (ebd., S. 19-26) alle drei Definitionselemente weiter aus: Das *methodische Erfassen* soll zum einen methodisch organisiert ablaufen, damit es überprüfbar ist. Dabei kommen klassische sozialwissenschaftliche Instrumente wie Fragebögen, Interviews, Tests, Checklisten oder auch Beobachtungsbögen zum Einsatz. Wichtig ist dabei, dass man nach der Erhebung weiß, wie die Daten zustande kamen, wobei die Aussage „Einige Teilnehmende wurden befragt“ laut Reischmann zu unpräzise sei, wohingegen die Aussage „Jeder fünfte der alphabetisch sortierten Teilnehmendenliste“ diesem Anspruch genügt. Zum anderen soll das Ergebnis der Datenerhebung auch dokumentiert werden, um den Anschein von Willkür oder subjektiven Interpretationen zu vermeiden. Das *begründete Bewerten* geschieht bei der Evaluation durch Vergleich von Ist-Werten mit (am besten vorab) definierten Soll-Werten. Für die Bewertung ist es zudem unabdingbar, sich für bestimmte Bewertungskriterien zu entscheiden und diese Entscheidung zu begründen. Denn um beispielsweise die Frage „War die Veranstaltung gut?“ zu beantworten, muss zunächst definiert werden, was unter „gut“ zu verstehen ist und je nach ausgewähltem Kriterium lässt sich diese Frage unterschiedlich beantworten. Nimmt man den Selbstlernanteil der Teilnehmenden als Kriterium, kann sich ein gänzlich anderes Bild zeigen, als wenn die Zufriedenheit der Teilnehmenden mit der Veranstaltung herangezogen wird. Die Klärung der Bewertungskriterien zu Beginn der Evaluation sollte dazu führen, dass die Erwartungen und Zielvorstellungen aller Beteiligten reflektiert werden (vgl. ebd., S. 39). Im dritten Definitionspunkt benennt Reischmann verschiedene Evaluationszwecke, die er als unterschiedliche didaktische Funktionen von Evaluation begreift, denn durch Evaluation soll Praxis „eingeschätzt, verstanden und verbessert werden“ (ebd., S. 24). Evaluation soll aufzeigen, was eine Veranstaltung bewirkt hat (*Wirkungskontrolle*), sie soll dazu beitragen, zukünftige Veranstaltungen besser zu gestalten (*didaktische Steuerung*) und Evaluation dient der *Reflexion* von Veranstaltungen.

Reischmanns Ausführungen zu den drei Definitionselementen kulminieren immer auch in der jeweiligen Festlegung von Ausschlusskriterien (vgl. ebd., S. 19-26). Nach seiner Ansicht sollte nicht von Evaluation gesprochen werden,

1. wenn keine Dokumentation vorliegt oder wenn Aussagen nicht mit zugänglichen Daten begründet werden können,
2. wenn nicht begründet bewertet, sondern nur die Teilnehmenden nach ihrer Wertung befragt werden oder nur exploriert wird im Sinne von „Wir schauen mal, was wir im Untersuchungsfeld so alles finden“ (auch wenn dabei umfangreiches Datenmaterial herangezogen wird),<sup>23</sup>
3. wenn es bei einer Untersuchung nur um allgemeine Erkenntnisse geht ohne direkte Konsequenzen bezüglich des Evaluationsgegenstandes.

Anders als bei Gerl, für den auch ein Seminarfeedback eine Evaluationsmethode darstellt, grenzt Reischmann Evaluation explizit von einem unsystematisch durchgeführten Seminarfeedback ab. Ein einfacher Seminarrückblick oder die Meinungsäußerung der Kursleiter zur Raumsituation seien noch keine Evaluation, sofern sie nicht methodisch kontrolliert stattfinden. Dennoch seien solche Formen für die Steuerung von Lernprozessen relevant. (Vgl. ebd., S. 30-31)

An dieser Stelle soll bewusst keine Festlegung auf eine dieser drei Definitionen für diese Arbeit erfolgen, denn erst bei der Durchführung der Analyse wird sich zeigen, welche dieser Evaluationsdefinitionen einen analytischen Wert hat. Eine Festlegung zum jetzigen Zeitpunkt hätte zudem große Konsequenzen: Eventuell wäre die Evaluationspraxis in den untersuchten Weiterbildungsorganisationen gar nicht als Evaluationspraxis zu bezeichnen. Stattdessen sollen die Definitionen nebeneinander stehen bleiben und als komplexe Analysefolie dienen, durch die ein Blick auf die Evaluationspraxis in Weiterbildungsorganisationen bzw. auf das, was die Organisationen darunter verstehen, geworfen wird.

### **2.3 Abgrenzung von Evaluation, Bildungscontrolling und Qualitätsmanagement**

Qualitätsmanagement, Evaluation und Bildungscontrolling stellen eine rotierende Begriffstrias dar: Je nachdem, wie weit man die einzelnen Begriffe fasst, erscheint mal der eine, dann der andere Terminus Oberbegriff zu sein und mal scheinen alle auf der gleichen Ebene zu liegen und bei genauerer Betrachtung stellen sich die Konzepte nicht nur ähnlich, sondern in Teilen sogar deckungsgleich dar. Als Grund

---

23 Unter Verweis auf den fünften Nützlichkeitsstandard der DeGEval führt Reischmann (2004, S. 42) an anderer Stelle aus: „Ohne explizite Bewertungskriterien und -maßstäbe keine Evaluation.“

für diese Begriffsvermischung bemerkt Schöni (2009, S. 56): „Die Ansätze sind aus unterschiedlichen Disziplinen hervorgegangen und in unterschiedlichen Anwendungsbereichen weiterentwickelt worden, und in der Praxis bestehen zahlreiche Überschneidungen. Zu strikte Definitionen laufen Gefahr, von der Praxis rasch überholt zu werden.“ Vor diesem Hintergrund erscheint es bezüglich Weiterbildungs-evaluation notwendig, diese gegenüber den konkurrierenden Konzepten Bildungscontrolling und Qualitätsmanagement abzugrenzen bzw. das Zusammenspiel dieser Konzepte zu erörtern. Das hier vorgestellte Verständnis beansprucht keineswegs Allgemeingültigkeit, sondern soll die drei Begriffe für die vorliegende Arbeit sinnvoll miteinander arrangieren.

Qualitätsmanagement (QM) wird in der Regel sehr breit definiert, wie etwa bei Veltjens (2010, S. 253-254) im Wörterbuch Erwachsenenbildung, wo es heißt: „Die Philosophie des QM kommt aus der Industrie und ist eine Bezeichnung für alle organisationsbezogenen Maßnahmen, die der Entwicklung und Verbesserung von Produkten und Dienstleistungen jeder Art dienen.“ Nach der internationalen Normenreihe DIN EN ISO 9000 versteht man unter Qualitätsmanagement „aufeinander abgestimmte Tätigkeiten zum Leiten und Lenken einer Organisation bezüglich Qualität“ (Brüggemann & Bremer, 2005, S. 122), worin zum Ausdruck kommt, dass das „managen“ von Qualität explizit als Führungsaufgabe betrachtet wird (vgl. z.B. Brunner & Wagner, 2008). Auch bei der Lernerorientierten Qualitätstestierung in der Weiterbildung handelt es sich um ein Qualitätsmanagementverfahren, wobei von den Entwicklern des Modells besonders der Aspekt der Qualitätsentwicklung betont wird (vgl. ausführlich hierzu Abschnitt 3.1). Zwar könnte man Verfahren zum Qualitätsmanagement, wie in Abschnitt 2.2 geschehen, auf der Mesoebene als Evaluationsverfahren verstehen, doch ist häufiger die nachvollziehbare Sichtweise anzutreffen, dass Evaluation im Qualitätsmanagement nur ein Instrument unter vielen darstellt.<sup>24</sup> Entsprechend dieser Unterordnung von Evaluation gegenüber dem Qualitätsmanagement existiert im LQW-Modell ein eigener Qualitätsbereich für die Evaluation der Bildungsprozesse. Evaluation ist jedoch nicht nur ein Teil von Qualitätsmanagement, denn wie Gutknecht-Gmeiner (2009, S. 13-7) treffend formuliert:

---

24 Gnahs (2010, S. 280) beginnt kennzeichnend für diese Sichtweise einen Satz mit folgenden Worten: „Qualitätsmanagement und – als Teilaspekt davon – Evaluierung ...“. Auch der Arbeitsstab Forum Bildung (2001, S. 7) äußert sich entsprechend: „Demnach umfasst Qualitätssicherung das aktive Bemühen darum, eine bestimmte Qualität zu erreichen, zu halten oder zu verbessern. Bei Evaluation handelt es sich dagegen um Aktivitäten, mit denen festgestellt werden soll, welche Qualität zu einem bestimmten Zeitpunkt erreicht worden ist. Qualitätssicherung ohne Evaluation ist nicht möglich.“

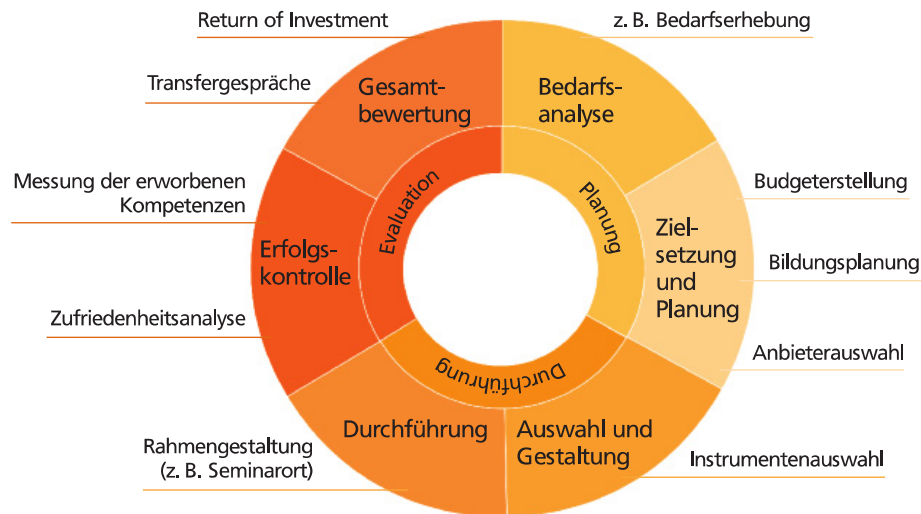
„Evaluation kann, muss aber nicht zwangsläufig, Qualitätsmanagement unterstützen.“

Einen anderen Fokus als Qualitätsmanagement nimmt das Konzept des Bildungscontrollings ein, dessen Ursprung und auch heutige Bedeutung in wirtschaftsorientierten Controlling-Konzepten zu finden ist und häufig mit der betrieblichen Bildung in Zusammenhang gebracht wird.<sup>25</sup> Eine praxisorientierte Definition legt Meier (2009, S. 13) vor: „Bildungscontrolling bezeichnet die Steuerung von Bildungsprozessen. Ziel hierbei ist, mit geeigneten Messgrößen und Kriterien zu überprüfen, ob die Bildungsarbeit methodisch effektiv und wirtschaftlich effizient vorgeht – von der Planung über die Durchführung bis zur Kontrolle der Ergebnisse.“ Wie das Qualitätsmanagement versteht sich auch Bildungscontrolling als „managementorientierter Steuerungsansatz“ (Schöni, 2009, S. 55). Konkret arbeitet Bildungscontrolling zumeist mit Soll-Ist-Vergleichen von betriebswirtschaftlichen Kennzahlen, ist also dezidiert an der ökonomischen Perspektive von Bildung interessiert. So weisen Beywl und Niestroj (2009, S. 22-23) in ihrem Evaluationsglossar darauf hin, dass im Mittelpunkt von Bildungscontrolling „die Bestimmung des organisationalen Wertschöpfungs- und Nutzenbeitrags des betreffenden Programms mit seinen Bildungsinhalten in Relation zu dem von ihm beanspruchten Aufwand“ steht. Das Tagungsprogramm des Fachkongresses zum Bildungscontrolling, der im Jahre 2010 bereits zum achten Mal in Deutschland stattfand, veranschaulicht zudem die Fokussierung des Bildungscontrollings auf den betrieblichen Weiterbildungsbereich. Zahlreiche Referentinnen und Referenten berichten über konkrete Erfahrungen mit Bildungscontrolling in großen Unternehmen oder thematisieren allgemein die Umsetzung im betrieblichen Kontext (vgl. [www.bildungs-controlling.com](http://www.bildungs-controlling.com)). Käßpflinger (2009, S. 2) begreift Bildungscontrolling als „ein Konzept, das ein Ablauf- und Phasenmodell für das Management betrieblicher Weiterbildung zur Verfügung stellt“ und greift zur Veranschaulichung auf eine kreisförmige Darstellung von Markowitsch (2009) zurück, die in Abb. 15 zu sehen ist. Evaluation wird in diesem Modell ausdrücklich als Teilbereich von Bildungscontrolling aufgefasst. Evaluation umfasst dabei nur ein Drittel des Gesamtprozesses und teilt sich in Erfolgskontrolle und Gesamtbewertung auf.<sup>26</sup>

---

25 Bei Krekel und Seusing (1999) finden sich Begriffsbestimmungen, empirische Untersuchungsergebnisse und Umsetzungsbeispiele zum Thema Bildungscontrolling im betrieblichen Kontext.

26 In den in der Abbildung dargestellten vier Evaluationsschritten „Zufriedenheitsanalyse“, „Kompetenzmessung“, „Transfergespräche“ und „Return of Investment“ schimmert das Evaluationsmodell

**Abb. 15: Kreislauf des Bildungscontrollings (Quelle: Markowitsch, 2009, nach Käßlinger, 2009, S. 2)**

Stockmann (2006, S. 75) sieht sowohl im Qualitätsmanagement, insbesondere in der Tradition des Total Quality Managements, als auch im Controlling eine kontinuierliche Aufgabe, während er Evaluation eher eine singuläre, periodisch wiederkehrende Aufgabe zuschreibt. Wie es auch in Abb. 15 zum Ausdruck kommt, ist Evaluation als Teilbereich von Bildungscontrolling ein Konzept, mit dem Daten und Empfehlungen generiert werden können, enthält jedoch keine „Entscheidungs- oder Managementkomponente“ (Stockmann, 2006, S. 95). Anders als Qualitätsmanagement und Bildungscontrolling liefert Evaluation also Daten fürs Management ist aber selbst keine Managementaufgabe und kann auch von externen Personen betrieben werden.

Schöni (2009, S. 54-57) beschreibt Bildungscontrolling im Vergleich zu Qualitätsmanagement und Evaluation als umfassendstes Konzept, das die *Gesamtsteuerung* zur Aufgabe habe, während prozessorientiertes Qualitätsmanagement vorwiegend nach der *Prozessbeherrschung* und Evaluation nur nach dem *Qualitätsniveau* frage. Der Qualitätsmanagementansatz von LQW versteht sich jedoch nicht nur als prozessorientiert, sondern ebenso als Mittel zur Steuerung einer gesamten Weiterbildungseinrichtung und sieht deshalb nicht nur einen eigenen Qualitätsbereich für Evaluation, sondern auch für Controlling vor, ordnet also gewissermaßen beide Begriffe als Unterbegriffe des Qualitätsmanagements an (vgl. Abschnitt 3.1).

Als pragmatisch orientiertes Fazit lässt sich für diese Arbeit folgendes festhalten:

---

von Kirkpatrick durch, das besonders im betrieblichen Kontext verbreitet ist und im Abschnitt 2.6.1 erläutert wird.

- Die Bedeutung, die Prominenz und die Verwendungshäufigkeit der Begriffe Qualitätsmanagement, Bildungscontrolling und Evaluation hängen stark von der Disziplin und dem Kontext ab, den man als Referenzrahmen wählt.
- Qualitätsmanagement, insbesondere im Verständnis nach LQW, stellt im Vergleich zu Bildungscontrolling und Evaluation das umfassendste Konzept dar.
- Bildungscontrolling wird zwar manchmal gegenüber Qualitätsmanagement als eigenständiges Verfahren verstanden, ist jedoch deutlich enger gefasst, hat eine stärker betriebswirtschaftliche und ökonomische Ausrichtung und ist tendenziell häufiger in der beruflichen, insbesondere betrieblichen Bildung anzutreffen.
- Während Qualitätsmanagement und auch Bildungscontrolling vornehmlich auf betriebswirtschaftliche Wurzeln zurückblicken kann, liegen die wissenschaftstheoretischen Wurzeln von Evaluation vorrangig in den Sozialwissenschaften (vgl. Stockmann, 2006, S. 86). Evaluation wird als Methode bzw. Instrument sowohl im Rahmen von Qualitätsmanagement als auch im Rahmen von Bildungscontrolling angewendet.

## 2.4 Unterscheidungsmerkmale für Evaluationen von Weiterbildungsprozessen

Letztendlich ist jede Evaluation, auch in der Weiterbildung, durch folgende Frage gekennzeichnet: *Warum* (Zweck) wird *was* (Gegenstand) mit *welchem Fokus* (Fragestellung) *mit welchem Maßstab* (Bewertungskriterium) und *mit welchem Vorgehen* (Methodik) beschrieben und bewertet? Da in der Literatur über Evaluationen von Weiterbildungsprozessen nicht immer klare Grenzen zwischen den ersten vier Aspekten gezogen werden, sollen diese im Folgenden detailliert betrachtet und ihre gegenseitigen Verknüpfungen und Bezüge dargelegt werden. Die ausführliche Beschäftigung mit Evaluationsmethoden und -instrumenten in der Literatur wird in einem eigenen Abschnitt 2.5 aufgegriffen.

### 2.4.1 Zwecke: Warum wird evaluiert?

Die in Abschnitt 1.5 vorgestellten allgemeinen Evaluationszwecke Erkenntnis, Kontrolle, Entwicklung und Legitimation gelten letztendlich auch für Evaluationen von Weiterbildungsprozessen. So kann eine Evaluation einen Erkenntnisgewinn bezüglich der Teilnehmendenstruktur verfolgen, der Kontrolle von neu eingeführten Kursen dienen, die Entwicklung und vor allem die Optimierung neuer oder bestehender

Kurskonzepte vorantreiben oder Grundlage für eine Legitimation gegenüber Auftraggebern einer Bildungsmaßnahme darstellen. Die Frage nach dem „Warum“ von Evaluationen wurde im Bereich der Weiterbildung nicht nur auf allgemeiner Ebene, sondern von zahlreichen Autoren/innen sehr konkret auf den Gegenstand Weiterbildung bezogen beantwortet.

Aufgrund seines Verständnisses von Evaluation als Erhöhung des Reflexionsgrades von oder in Lernsituationen (siehe oben) benennt Gerl (1983, S. 21-22) ausdrücklich Funktionen (Zwecke) von Evaluation, in denen seine pädagogischen Vorstellungen von Evaluation zum Tragen kommen. An erster Stelle hat Evaluation für ihn eine *didaktische Funktion*, wenn sie eingebettet in einem Lernprozess stattfindet, so dass die Resultate direkt auf diesen zurückwirken – Scriven würde in diesem Zusammenhang von formativer Evaluation sprechen. „Evaluation prüft so gesehen nicht nur Lernergebnisse oder macht Lernprozesse transparent, sie ermöglicht und fördert auch bestimmte Ergebnisse, leitet selbst bestimmte Lernprozesse ein“ (Gerl, 1983, S. 22). Dieser Funktion ordnet Gerl weitere konkrete Funktionen unter, nämlich die Erweiterung der Mitwirkungsmöglichkeiten der Beteiligten, die Selbstkontrolle der Lernfortschritte und die Steigerung der Lernmotivation. Zweitens übernimmt Evaluation für Gerl eine *pädagogische Funktion*, indem über den Umweg pädagogischer Theoriebildung die Evaluation zeitversetzt auf zukünftige Lernprozesse zurückwirkt. Diese Funktion lässt sich demnach als summativ charakterisieren.

Will und Blickhan (1987, S. 50-51) diskutieren die didaktische Funktion von Evaluation unter dem Titel „Evaluation als Intervention“ und benennen folgende Ziele intervenierender und bildungsunterstützender Evaluation. Eine intervenierende Evaluation ...

- bindet Dozenten stärker in das Treatment ein,
- erlaubt es, Inhalte, Methoden und Ablauforganisation ad hoc an die Bedürfnisse der Teilnehmenden anzupassen,
- gibt den Teilnehmenden Rückmeldung über ihren Lernstand, unterstützt ihren Lernprozess, erhöht ihre Bereitschaft und Fähigkeit zur Reflexion,
- zielt darauf, den Transferprozess zu unterstützen und die Wirkung des Treatments zu optimieren,
- kann Zusammenhalt und Kooperation unter den Lernenden, den Lehrenden und Lernenden sowie den Lehrenden und den Organisationen festigen.

Auch Reischmann, der Evaluation ausdrücklich als andragogisches Handeln und „Teil der andragogisch-didaktischen Planung und Durchführung“ (Reischmann, 2006, S. 263) begreift, betont die didaktische Funktion, die Evaluation in der Weiterbildung zukommt. Konkret sollen Evaluationen zurückblickend kontrollieren, was Bildungsmaßnahmen bewirkt haben; sie sollen steuern, indem sie vorausschauend dabei helfen, zukünftige Maßnahmen besser zu gestalten; und sie dienen der Reflektion und dem Verständnis von Bildungsmaßnahmen (vgl. ebd., S. 24). Die Evaluationsfunktionen bei Reischmann beziehen sich demnach primär auf die Verbesserung zukünftiger Maßnahmen, während Gerls Funktionen die Einbeziehung der Teilnehmenden in die Evaluation und die Erhöhung von deren Lernen durch die Integration der Evaluation in den Lehr-Lern-Prozess fokussieren. Aufbauend auf einem Verständnis von selbständigen und mündigen erwachsenen Lernenden benennt Reischmann weitere, für die Weiterbildung spezifische, Evaluationsfunktionen (vgl. ebd. S. 88-89). Demnach sollen mithilfe von Evaluationen Bildungsangebote auf ihre Qualität hin überprüft werden, und nicht Lernende hinsichtlich ihrer Leistungen verglichen werden. Formative Evaluationen sollen Lernhilfe sein, sind Teil des Lernprozesses und dienen dem Lernerfolg. Evaluation unterstützt die Eigensteuerung des Lernens durch vorwiegend Selbstkontrolle, aber auch Fremdkontrolle.

Zech (2008, S. 88-89) führt drei Hauptzwecke von Evaluation an. Die *Kontrolle von durchgeführten Bildungsmaßnahmen* lasse im Nachhinein die Zufriedenheit der Teilnehmenden, das Stattfinden eines Lernprozesses und den Beitrag von Räumlichkeiten, des Lehrverhaltens und anderer Faktoren für den Lernerfolg erkennbar werden. Die *Steuerung von Bildungsangeboten* umfasst wie bei Reischmann die Verbesserung zukünftiger Kurse, allerdings bei Zech konkretisiert durch den gleichzeitigen Abbau von Defiziten und den Ausbau vorhandener Stärken. Und drittens dient auch für Zech die Evaluation der Reflexion von Lernerfolg und Lernprozess:

„Durch eine begleitende oder abschließende Befragung (formative oder summative Evaluation) erhalten viele Teilnehmende erst die Gelegenheit, zur Besinnung und zur Klärung des eigenen Lernerfolges. So kann z.B. ein Fragebogen auf die eigenen Empfindungen und Entwicklungen aufmerksam machen, die ansonsten möglicherweise überhaupt nicht beachtet worden wären. Letztlich befähigt eine Evaluation Lernende, Lehrende und Auftraggeber auch in späteren Lernsituationen kompetente, kundige und kritische Einschätzungen zu erbrin-



gen. Sie schult den Blick für Bildungsangebote und Situationen, die gelingendes Lernen ermöglichen.“ (Zech, 2008, S. 89)

Wenngleich in der Literatur zur Weiterbildungsevaluation Zwecke dominieren, welche vor allem den Organisationen nutzen, so sollten doch die bisherigen Darstellungen die mögliche didaktische Funktion im Sinne einer Lernunterstützung und Lernintervention verdeutlichen haben. Neben den bisher genannten Evaluationszwecken lassen sich zahlreiche weitere benennen. So dient Evaluation der Sicherung von Qualität sowie als Planungs- und Steuerungsinstrument (vgl. Kuper, 2008; Wesseler, 2010). Darüber hinaus kann Evaluation eine kommunikationsfördernde Funktion übernehmen und zur Team- und Organisationsentwicklung beitragen, wenn viele der Beteiligten an der Planung, Durchführung und Auswertung der Evaluation einbezogen werden und dadurch eine gemeinsame Gesprächs-, Bewertungs- und Handlungsbasis geschaffen wird (vgl. Will, Winteler & Krapp, 1987a, S. 25). Evaluation in der Weiterbildung kann auch den (unbeabsichtigten) Nebeneffekt hervorrufen, dass sich die Lehrenden und Programmverantwortlichen darüber verständigen, welche Zielsetzungen mit einem Seminar verfolgt werden. Evaluation kann also zur Herausarbeitung und Klärung von Lernzielen eingesetzt werden.

Zusammenfassend lassen sich folgende, nicht ganz trennscharfe Evaluationszwecke im Bereich der Weiterbildung differenzieren, wobei sich die Zwecke danach unterscheiden, inwieweit sie sich auf den Gegenstand „Bildungsprozess“, die Lernenden oder die Evaluierenden konzentrieren. Evaluation in der Weiterbildung dient ...

- der Wissens- und Erkenntnisgenerierung,
- der (Erfolgs-)Kontrolle und Zielüberprüfung,
- der Entscheidungsfindung und Steuerung,
- der Verbesserung, Optimierung und Entwicklung,
- der Rechenschaftslegung und Legitimation,
- der Unterstützung von Didaktik und Lernprozess,
- der Reflexionsförderung bei Teilnehmenden und Durchführenden,
- der Kommunikationsförderung, Team- und Organisationsentwicklung und
- der Weiterbildung der Durchführenden.

### 2.4.2 Gegenstände: Was wird evaluiert?

Im Rahmen einer Evaluation wird üblicherweise das, was evaluiert wird, also das, was bewertet werden soll und worüber Konsequenzen getroffen werden, als Gegenstand bezeichnet. Diese Evaluationsgegenstände können unterschiedlich komplex sein, und zwar sowohl hinsichtlich ihres Umfangs als auch ihres Inhalts und der Anzahl der Stakeholder und vielem mehr. Es stellt offensichtlich einen Unterschied dar, ob ein Vortrag, ein Abendkurs mit zehn Terminen, alle Kurse eines Fachbereichs oder eine dreimonatige Maßnahme für arbeitslose Jugendliche evaluiert wird. Evaluationsmodelle, wie das CIPP-Modell von Stufflebeam und der Programmbaum von Beywl (vgl. Abschnitt 2.6) helfen dabei, einen komplexen Gegenstand in seiner Gänze zu erfassen, zu analysieren und zu beschreiben und auch Teilaspekte für die Evaluierung auszuwählen.

Die Beschreibungen von Evaluationsgegenständen in der Weiterbildung sind in der Literatur sehr uneinheitlich und folgen häufig unterschiedlichen Logiken und Herangehensweisen. Wesseler (2010, S. 1035-1036) verweist in diesem Zusammenhang auf das von Windham (1988) leicht modifizierte CIPP-Modell von Stufflebeam, das vereinfacht dargestellt zwischen Context, Input, Process, Output und Outcome differenziert. Zum Gegenstand „Input“ zählen beispielsweise die vorhandenen Qualifikationen der Teilnehmenden und Kompetenzen der Lehrenden und zum „Process“ die didaktischen Medien, aber auch die Beziehungen zwischen Lernenden und Lehrenden.

Häufig finden sich auch Auflistungen an konkreten Gegenständen, die sich in der Weiterbildung evaluieren lassen, ohne eine Gruppierung oder Zuordnung vorzunehmen. So führt Zech (2008, S. 89) folgende Beispiele für in der Weiterbildung evaluierbare Gegenstände an:

- die organisationalen Rahmenbedingungen,
- den begleitend angebotenen Service,
- die Arbeitsleistung der planend und administrativ Beschäftigten,
- die Leistungen der Lehrenden,
- die Zufriedenheit der Teilnehmenden,
- die Zufriedenheit der Auftraggeber,
- die Lernerfolge der Teilnehmenden,

- den Transfer des Gelernten in die Praxis,
- die erreichten Abschlüsse und Zertifikate,
- die Wiedereingliederung in den Beruf.

Und auch Reischmann (2006, S. 37) gibt ähnliche Beispiele dafür, was man evaluieren kann, nämlich Programmhefte, Teilnahme, Kurserfolg, Zielgruppenerreichung, Inhalte, Kosten, Nutzen, Schwachstellen. Bezüglich möglicher Bedenken von Kursleitenden gegenüber Evaluation betont Reischmann (2006, S. 33) zudem, dass bei der Evaluation nicht „der Teilnehmer, Kursleiter, Planer“ auf dem Prüfstand stehe, sondern die pädagogisch/andragogische Maßnahme und deren Zielerreichung. Meiner Einschätzung nach darf diese Anmerkung jedoch nicht darüber hinwegtäuschen, dass Evaluationen häufig auch zur Verbesserung des Kursleiterhandelns beitragen sollen – und müssen, weil eben hierin ein wesentliches Potenzial zur Optimierung von Lehr-Lern-Prozessen liegt.

Setzt man die Überlegungen zum Weiterbildungsprozess aus Abschnitt 2.1 fort, so stellen sowohl Abb. 11 als auch Abb. 10 einen geeigneten Rahmen für die Eingrenzung des evaluierten Gegenstandes dar: Zunächst lässt sich bestimmen, welcher Schritt des Bildungsprozesses im Fokus der Evaluation stehen soll, also ob beispielsweise der Erstkontakt, die Anmeldung, die Bildungsveranstaltung selbst oder der Transfer evaluiert werden sollen. Mit dieser Bestimmung geht dann eine – oder gehen dann auch mehrere – Dimensionen des Bildungsprozesses einher, und man kann sich bewusst machen, ob man auf der Ebene der Lernenden, der Lehrenden, der Lerninfrastruktur, der Organisation oder des Systemumfelds oder einer Kombination der Ebenen evaluieren möchte.

Für die weitere Betrachtung, insbesondere bezüglich der Bewertungskriterien, ist wichtig festzuhalten, dass die Benennung von Evaluationsgegenständen auf sehr unterschiedlichem Abstraktionsniveau oder besser ausgedrückt: Konkretionsniveau geschehen kann, so ist der Gegenstand „die Bildungsmaßnahme“ deutlich abstrakter als der Gegenstand „das Kursleiterhandeln“. Zudem erscheint es relevant darauf hinzuweisen, dass es nur Sinn hat etwas zu evaluieren, das sich auch verändern oder umgestalten lässt, zumindest wenn die Evaluation eine Verbesserung und Optimierung verfolgt. Beispielsweise besteht häufig ein sehr eingeschränkter Spielraum, was die Räume einer Organisation und deren Ausstattung betrifft, dennoch wird häufig die Raumsituation evaluiert.

### 2.4.3 Fragestellungen: Mit welchem Fokus wird evaluiert?

Fragestellungen liefern einer Evaluation die notwendige Richtung, sie helfen dabei, gezielt Daten zu erheben und die Ergebnisnutzung bereits zu Beginn der Evaluation zu antizipieren. Die Fragestellungen nehmen also in der Evaluation eine sehr prominente Rolle ein, und umso wichtiger ist es, hilfreiche und nutzbringende Fragen zu stellen: „Die beste Evaluation nützt nichts, wenn sie die falschen Fragen stellt“ (Wessler, 2010, S. 1042). Für Fragestellungen lassen sich mehrere Qualitätsaspekte definieren. Nach Reischmann (2006, S. 37) sollten sie nicht allumfassend, sondern eingegrenzt und klar sein. Fragen wie „Ist unser Programm gut?“ oder „Sind unsere Kursleiter ihr Geld wert?“ seien aufgrund ihrer Breite nicht evaluierbar. Beywl et al. (2007, S. 28) schlagen zur Generierung und Bewertung der Eignung von Fragestellungen ein Raster mit dem Akronym „f.ö.r.d.e.r.n“ vor:

- Fokussiert – es ist klar, auf welchen Bestandteil des Evaluationsgegenstandes sich die Fragestellung richtet.
- Öffnend und gradierend – die Fragestellung lässt sich nicht mit ja/nein beantworten, sondern es wird nach Ausmaßen, Ausprägungen gefragt.
- Realistisch – die Fragestellung bezieht sich auf Sachverhalte, die tatsächlich auf den Evaluationsgegenstand zurückzuführen sind.
- Deutlich – die Fragestellung ist verständlich formuliert und enthält keine unklaren Begriffe.
- Empirisch beantwortbar – die Fragestellung lässt sich mittels Daten beantworten und die Sachverhalte sind operationalisierbar.
- Ressourcenangepasst – die Fragestellung berücksichtigt vorhandene Mittel und Zeitbudgets.
- Nützlich – die Fragestellung bezieht sich auf das Informationsinteresse der Stakeholder, welche die Evaluationsergebnisse nutzen werden.

Um evaluierbare Fragestellungen zu finden, bedarf es der Erfahrung und Kompetenz, die Evaluierende mitbringen müssen, nichtsdestotrotz lassen sich einige Denkanstöße für Fragestellungen im Kontext von Weiterbildung geben (vgl. Beywl et al., 2007, S. 28): Wo waren Unsicherheiten bei der Konzepterstellung und Planung? Wo sind kritische Situationen bei der Durchführung zu erwarten? Welche Zielerreichung ist für den Erfolg des Evaluationsgegenstands zentral? Mit welchen Ergebnissen kann am deutlichsten gezeigt werden, dass das Programm zweckmäßig war? Fol-

gende Beispiele (angelehnt an Rädiker, 2009, S. 135) verdeutlichen konkrete und enggeführte, gut geeignete Fragestellungen einer Evaluation von Weiterbildungsprozessen und zeigen, dass sich die Fragestellungen direkt auf dazugehörigen Gegenstände beziehen:

- Organisationale Rahmenbedingungen: Inwieweit wirkt sich die neue Möblierung auf die Zufriedenheit der Teilnehmenden aus?
- Einzelne Kurse oder Kursreihen: Inwieweit entsprechen die Kursinhalte den Erwartungen? Wie groß sind die Lernerfolge in Bezug auf xy? Welche Probleme haben die Teilnehmenden beim Transfer des Gelernten in die Praxis?
- Lernmaterialien: Sollten wir besser CD-ROM A oder B anschaffen?
- Zielgruppenerreichung: Inwieweit erreichen wir die Teilnehmenden, die wir laut unseres Leitbildes erreichen wollen?
- Ankündigungstexte im Programm: Inwieweit fühlen sich unsere Zielgruppen von den Formulierungen angesprochen?

Neben dem Zusammenhang von Fragestellungen und Gegenständen ist es abschließend auch wichtig, die Beziehung zu den Zwecken hervorzuheben: Die Fragestellungen dienen dazu, einen konkreten Evaluationszweck zu erfüllen, und dies gilt es bei der Formulierung von Fragestellungen zu berücksichtigen.

#### **2.4.4 Bewertungskriterien: Wie wird bewertet?**

Evaluation umfasst immer den Aspekt der Bewertung. Wenn ein Gegenstand bewertet werden soll, setzt dies folglich Kriterien voraus, nach denen die Bewertung vonstattengehen soll. Ein Kriterium wird üblicherweise als Beurteilungsdimension oder Beurteilungsgesichtspunkt definiert. Hierbei ist zu beachten, dass Kriterien einen unterschiedlichen Konkretionsgrad aufweisen können. Als ganz allgemeine, sogenannte generische Kriterien, die für alle Gegenstände Gültigkeit haben, gelten die Güte (intrinsischer Wert), die Tauglichkeit (Gebrauchswert) und die Bedeutsamkeit (Wichtigkeit) eines Gegenstandes (vgl. u.a. Beywl & Niestroj, 2009, S. 66; Scriven, 2007). Auch auf etwas konkreterer Ebene lassen sich weitere allgemeine Bewertungskriterien benennen. Stockmann (2006, S. 165) hat in der Praxis häufig verwendete Kriterien zusammengestellt. Hierzu zählen vor allem Output, Impact, Effek-

tivität, Outcome, Wirksamkeit und Effizienz.<sup>27</sup> Ergänzen lassen sich noch aus eher theoretischer Perspektive die Zielakzeptanz, die Zielgruppenrelevanz, der Zielgruppennutzen, die Zielgruppenerreichung, die ökologische Verträglichkeit und die Nachhaltigkeit (vgl. Stockmann, 2006, S. 166; Gollwitzer & Jäger, 2007, S. 7).

Für die Evaluation von Weiterbildungsprozessen im Rahmen des LQW-Modells führt Zech (2006, S. 68) den Lernerfolg, die Zufriedenheit der Auftraggebenden und die Einlösung des eigenen institutionellen Anspruchs als maßstabbildend, also als Kriterien für die Bewertung des Bildungsprozesses, an. Diese Auflistung wirft die Frage auf, in welchem Zusammenhang die evaluierten Gegenstände und die Kriterien stehen, da beispielsweise der Lernerfolg weiter oben in dieser Arbeit zugleich auch als Gegenstand bezeichnet wurde. Bei der Analyse einiger Selbstbeschreibungen von der Evaluationspraxis in Weiterbildungsorganisationen konnte festgestellt werden, dass die Abgrenzung von Bewertungskriterium und Evaluationsgegenstand Schwierigkeiten bereitet. Dies liegt darin begründet, dass die Organisationen teilweise sehr konkrete Evaluationsgegenstände benennen, die man auch als Bewertungskriterium auffassen könnte (vgl. Rädiker, 2010c, S. 162). Hierzu lässt sich folgende Beziehung festhalten: Je konkreter ein Gegenstand benannt wird, desto eher geht er in ein Bewertungskriterium über – es handelt sich gewissermaßen um eine Kette von Operationalisierungen, die man durchlaufen kann: „Beispielsweise kann die Qualität des Evaluationsgegenstandes Bildungsmaßnahme am Bewertungskriterium Teilnehmendenzufriedenheit abgelesen werden. Die Teilnehmendenzufriedenheit könnte jedoch auch selbst Evaluationsgegenstand sein und anhand der Kriterien Zufriedenheit mit Dozent, mit Kursmaterial, Programmankündigung und Räumlichkeiten erhoben werden. Für die Zufriedenheit mit Dozenten ließen sich wiederum Kriterien wie methodische Vielseitigkeit angeben usw.“ (ebd., S. 162). In Abb. 16 wurde dieser Zusammenhang anhand des Bildungsprozesses visualisiert. Für alle vier Stationen des Bildungsprozesses, vom Erstkontakt bis zum Transfer, können Bewertungskriterien angegeben werden, die ihrerseits selbst als Gegenstand fungieren können. Die Konkretisierungen liegen im Beispiel allerdings nicht auf dem gleichen Abstraktionsniveau.

---

27 Die Begriffe werden zum Teil im Rahmen des Evaluationsmodells „Programmbaum“ im Abschnitt 2.6.3 erläutert.

**Abb. 16: Das Zusammenspiel von Evaluationsgegenständen und Bewertungskriterien (eigene Darstellung)**

Bildungsprozess	Erstkontakt	Anmeldung	Bildungs- veranstaltung	Transfer
1. Konkretisierungs- schritt	Programmheft Webseite ...	Beratung ...	Inhalte Lehrhandeln Rahmenbedingun- gen ...	Übertragung des Gelernten in die Praxis ...
2. Konkretisierungs- schritt	Texte Layout Zugänglichkeit ...	Zufriedenheit mit Zielorientierung der Beratung Telefonische Erreichbarkeit ...	Nachvollzieh- barkeit der Inhalte Lernmaterialien, Selbstlernphasen Kurszeiten ...	Veränderte Einstellungen Optimiertes Handeln ...
..				

Um einen Gegenstand anhand eines Kriteriums zu bewerten, muss dieses Kriterium in beobachtbare und messbare Indikatoren<sup>28</sup> operationalisiert werden, sofern die Kriterien nicht direkt beobachtbar sind – was selten der Fall ist. Woran lässt sich beispielsweise ablesen, dass eine Teilnehmerin etwas gelernt hat oder ein Teilnehmer die Eingangsberatung gut fand? Wenn es beispielsweise darum geht, konkret beobachtbares Verhalten in Lernsituationen zu erfassen, empfiehlt Reischmann (2006, S. 185) unter Bezugnahme auf die Lernzielanalyse folgenden Wörtern den Vorzug zu geben, da sie wenig Interpretationsspielraum zulassen: wissen, verstehen, wirklich verstehen, zu würdigen wissen, die Bedeutung erfassen, glauben, können.

Durch die Operationalisierung entsteht eine Bewertungsskala, die als Maßstab für die Beurteilung des Gegenstandes herangezogen wird (vgl. Beywl, 2010b). Insbesondere wenn Evaluationen zur Überprüfung der Zielerreichung eingesetzt werden, ist es notwendig, auf diesem Maßstab Soll-Zustände oder Soll-Bereiche zu definieren, die eine Erfolgszone markieren. Soll-Werte können aus verschiedenen Überlegungen abgeleitet werden, beispielsweise aus bisherigen Evaluationsergebnissen des gleichen Bildungsangebots, vergleichbarer eigener Bildungsangebote oder vergleichbarer Bildungsangebote von Mitbewerbern.<sup>29</sup> Gerade in Organisationen, die das LQW-Modell anwenden, lassen sich Bewertungskriterien, Soll-Zustände und Zie-

28 Zu Indikatoren vgl. Rädiker (2009); Stockmann (2006, S. 217-224); Schnell, Hill und Esser (2008, S. 192ff.); Kromrey (2006).

29 In Anlehnung an Vedung (1999, S. 224), der für den öffentlichen Sektor zehn Bezugspunkte für Bewertungen nennt, lassen sich für Weiterbildung beispielsweise folgende Referenzpunkte nennen: vormals erreichte Werte, intra-/interorganisationaler Vergleich, beste empirische Praxis, Ziele des Programms, Erwartung der Teilnehmenden, Erwartung der Adressaten, Standards der Profession, Mindestansprüche, optimal vorstellbares Modell.

le aus den Ansprüchen des Leitbildes, der Definition gelungenen Lernens (vgl. Abschnitt 3.1), den allgemeinen Unternehmenszielen, den strategischen Entwicklungszielen einer Organisation, den Lernzielen und Konzepten von Veranstaltungen oder den Vorgaben von Auftraggebern deduzieren.

Die Auswahl von Bewertungskriterien ist predeterminiert durch die Evaluationszwecke, denn andere Zwecke führen zwangsläufig zu anderen Kriterien. Wenn eine Evaluation vornehmlich der wirtschaftlichen Legitimation dient, wird die Kosten-Nutzen-Analyse relevanter sein als die Zufriedenheit der Teilnehmenden mit dem Lernmaterial. In der Evaluationsliteratur wird darüber diskutiert, wer überhaupt die Bewertungskriterien auswählt und inwieweit die Beteiligten und Betroffenen an diesem Auswahlprozess partizipieren (vgl. Beywl et al., 2007; Stockmann, 2010c, S. 79-80 u.a.). In jedem Fall ist eine Einigung über die Kriterien zu Beginn der Evaluation zu empfehlen, weil ansonsten jede Wirkung des Evaluationsgegenstandes ad absurdum geführt werden kann. So können zwar beispielsweise die subjektive Zufriedenheit und die Bereicherung der Persönlichkeit hoch sein, aber der objektive Lernerfolg und der Nutzen für die Öffentlichkeit minimal (vgl. Abschnitt 2.7 zu den Herausforderungen bei der Durchführung von Evaluationen).

## **2.5 Methoden, Instrumente und Verfahren der Weiterbildungsevaluation**

Unter Evaluationsmethoden und -instrumenten sollen in dieser Arbeit alle Verfahren verstanden werden, die der Erhebung und Analyse von relevanten Daten für die Evaluation dienen.<sup>30</sup> Da es keine einheitliche Abgrenzung der Begriffe Evaluationsmethode, -instrument und -verfahren gibt, werden diese im Folgenden meist synonym verwendet. Gleich, welchen Begriff man bevorzugt, ist klar, dass die eingesetzten Verfahren eine besondere und wichtige Bedeutung für den Evaluationsprozess und die Evaluationsergebnisse haben, denn „Evaluationsverfahren sind wie Fischernetze: Man bekommt nur zu sehen, was die Maschen des Netzes zu ‚fangen‘ vermögen“ (Wesseler, 2010, S. 1043). Aus diesem Grund fordert der Standard G5 der DeGEval, dass sich eingesetzte Datenerhebungsverfahren Gütekriterien stellen müssen, welche sich an den Gütekriterien qualitativer und quantitativer Sozialforschung orientieren (vgl. DeGEval, 2008, S. 35). Dementsprechend finden sich in vielen Evaluationsbüchern zum Thema Methoden meist zusammenfassende oder überblicks-

---

30 Zum Teil wird der Begriff der Evaluationsmethoden in der Literatur deutlich breiter definiert und umfasst neben der Datenerhebung unter anderem auch „Methoden“ zur Steuerung, Auswertung und Berichterstattung (z.B. bei Beywl & Niestroj, 2009, S. 40).



artige Darstellungen von Methoden empirischer Sozialforschung (z.B. bei Meyer, 2007), die mehr oder weniger den Kontext der Evaluation berücksichtigen.

In der Fülle an Literatur über Evaluationsmethoden im Kontext der Weiterbildung findet sich nicht immer ein deutlich formulierter Bezug auf die Methoden empirischer Sozialforschung, wenngleich viele Autoren/innen das Wissen der Sozialwissenschaften über Methoden heranziehen. Sehr häufig werden zahlreiche einzelne Evaluationsmethoden vorgestellt oder mehr oder weniger systematisch aufgezählt. Diese reichen vom klassischen Fragebogen und Interviews über Hospitationen und Beobachtungen bis hin zu Lerntagebuch, Blitzlicht, Stimmungsbarometer, 360°-Feedback und bildlicher Ausdruck.<sup>31</sup> Die schillernde Vielfalt erweckt in gewissem Maße den Eindruck eines wenig sortierten Gemischtwarenladens, aus dem sich Evaluierende bedienen können. Zur systematischen Darstellung der Methoden wird im Folgenden auf die in der empirischen Sozialforschung übliche Unterteilung in *Befragung*, *Beobachtung* und *Materialanalysen* (vgl. z.B. Flick, 2009) zurückgegriffen und um das Evaluationsverfahren *Test/Lernkontrolle* ergänzt, das sich nicht direkt den drei Bereichen zuordnen lässt, aber häufig diskutiert wird.

### 2.5.1 Befragung

Zur Methode der Befragung zählen insbesondere der Fragebogen, das Einzelinterview und das Gruppeninterview, meist Gruppendiskussion genannt. Befragungen eignen sich insbesondere für die Erfassung von Meinungen und Einstellungen und sind das Mittel der Wahl, „wenn man Verhalten, Gefühle oder die Art und Weise wie Menschen ihre Welt sehen und interpretieren, nicht direkt beobachten kann“ (Reischmann, 2006, S. 137). Methoden der Befragung lassen sich hinsichtlich folgender Aspekte unterscheiden (vgl. u.a. Diekmann, 2007, S. 437):

- Art und Medium der Kommunikation: schriftlich, mündlich, online, telefonisch, persönlich;
- Ausmaß der Standardisierung der gestellten Fragen, der Antwortmöglichkeiten und der Reihenfolge der Fragen: gering bis hoch;
- Anzahl der Befragten: eine Person versus mehrere.

---

31 Berücksichtigt wurden bei dieser Aufzählung unter anderem folgende Quellen: Beywl et al. (2007, S. 33-34); Frankenreiter (1996, S. 39-47); Gerl (1983, S. 25-89); Götz (2001, S. 113-121); Häring (2003, S. 76-79); Reischmann (2006, S. 137-157); Will et al. (1987b); Zech (2008).

Standardisierte Fragebögen als schriftliche Form der Befragung gehören zu den am häufigsten eingesetzten Verfahren in der Evaluation. Interviews als klassische Form der mündlichen Befragung können sowohl persönlich als auch telefonisch mit Einzelpersonen durchgeführt werden.

Mit den einzelnen Formen der Befragung sind entsprechende Vor- und Nachteile verbunden. Standardisierte Fragebögen sind vor allem deshalb so beliebt, weil sie „schnell, billig und bei fast allen Fragestellungen anwendbar“ (Reischmann, 2006, S. 137) sind. Die schriftliche Form ermöglicht die Sicherstellung von Anonymität und kann daher zu ehrlicheren Antworten führen als ein persönliches Interview. Darüber hinaus ist die Belastung der Teilnehmenden im Sinne des Standards D1 „Angemessene Verfahren“ (vgl. DeGEval, 2008, S. 26) meist gering und der hohe Grad an Strukturierung und Standardisierung vereinfacht die Auswertung und ermöglicht es, verschiedene Befragtengruppen zu vergleichen. Standardisierte Befragungsinstrumente sind allerdings nur dann sinnvoll einsetzbar, wenn die möglichen Antworten der Befragten bekannt sind oder antizipiert werden können, denn geschlossene und hoch strukturierte Fragen rekurren auf die Problemsicht der Evaluierenden, während offene Fragen Antworten aus Sicht der Befragten hervorbringen (vgl. Reischmann, 2006, S. 139-140). Zu den Nachteilen einer standardisierten, schriftlichen Befragung zählt neben der meist fehlenden Nachfragemöglichkeit für die Befragten auch die Tatsache, dass die Evaluierenden meist nichts über die Kontexte und Beweggründe erfahren, die einen Teilnehmenden zu einer Antwort, meistens einem Kreuz auf einer Skala, veranlassen haben. Die Evaluierenden wissen dann beispielsweise, dass jemandem der Kurs zu langatmig war, aber nicht warum und in welchem Zusammenhang. Mit standardisierten Fragen lassen sich höchstens vorhandene und bewusste Verbesserungsvorschläge hinsichtlich Wichtigkeit und Präferenz einschätzen; um neue Vorschläge hervorzubringen, sind jedoch offene Frageformen ohne Antwortvorgaben notwendig, die allerdings meist aufwändiger auszuwerten sind.

Interviews mit Teilnehmenden oder anderen Datengebern/innen in der Evaluation verschaffen den Befragten Gehör und erlauben subjektive Sichtweisen zu ermitteln und reduzieren darüber hinaus Interpretationsspielräume, weil die Befragten zu ihren Einschätzungen auch Begründungen und Kontextinformationen liefern (vgl. Kuckartz et al., 2008, S. 67). Interviews werden selten abgebrochen und erlauben Nachfragen bei Unklarheiten (vgl. Reischmann, 2006, S. 142), sie sind allerdings im

Verhältnis zu Fragebögen aufwändiger was die Durchführung als auch die Auswertung betrifft.

Alle Befragungen sind reaktiv, das heißt die Befragten wissen, dass eine Datenerhebung für eine Evaluation stattfindet und die Evaluation selbst kann den evaluierten Gegenstand beeinflussen und verändern. Dies kann zu Verzerrungen der Antworten durch soziale Erwünschtheit, durch Überlegungen der Befragten, welches Ziel die Evaluation verfolgt, und anderen Beeinträchtigungen führen, die allesamt die Validität der Daten senken (vgl. Diekmann, 2007, S. 446-470; Gollwitzer & Jäger, 2007, S. 147-149).

Dieser Umstand kann für die Forschung sehr gravierend sein, da diese an unverfälschten Daten interessiert ist. Im Rahmen der Evaluation in der Weiterbildung kann die Reaktivität der Methoden jedoch konstruktiv genutzt werden, wenn eine Evaluation bewusst einen lernunterstützenden und -motivierenden Zweck verfolgt und Lernenden beispielsweise ein Fragebogen zur Reflexion des eigenen Lernprozesses vorgelegt wird.

### **2.5.2 Beobachtung**

Beobachtungen im Rahmen von Evaluationen nehmen soziale Prozesse und Gegebenheiten in den Blick (vgl. Beywl et al., 2007, S. 33). Zur Methode der Beobachtung gehören insbesondere die Hospitation von Bildungsveranstaltungen, die Teilnahme an Beratungsgesprächen und die Beobachtung von Teilnehmenden in ihrem Arbeitsumfeld. Ebenso wie die Berichte von Kursleitenden Daten aus eigenen Beobachtungen enthalten können, zählt auch die Begehung einer Bildungsstätte, um die Räumlichkeiten zu begutachten, zu den Beobachtungsmethoden. Reischmann (2006, S. 143) weist darauf hin, dass Beobachtungen zu geeigneten Datenerhebungsmethoden für die Evaluation werden, wenn sie bestimmten Kriterien genügen, die er von Merriam (1988) übernimmt: Beobachtungen müssen einem Zweck dienen, systematisch geplant sein, aufgezeichnet und dokumentiert sowie auf ihre Aussagequalität kontrolliert werden. Bortz und Döring (2006, S. 263) setzen darüber hinaus für systematische Beobachtungen einen Beobachtungsplan voraus, der unter anderem festschreibt, was zu beobachten und was für die Beobachtung unwesentlich ist, ob bzw. in welcher Weise das Beobachtete gedeutet werden darf und wann und wo die Beobachtung stattfindet.

Unterschiedliche Formen der Beobachtung lassen sich nach verschiedenen Kriterien systematisieren (vgl. Bortz & Döring, 2006, S. 266-268; Diekmann, 2007, S. 563-569):

- *Teilnehmend vs. nicht-teilnehmend*: Ist die beobachtende Person Teil des beobachteten Geschehens oder nicht? Je stärker eine Person in das zu beobachtende Geschehen involviert ist, desto stärker beeinflusst sie dieses und desto schwieriger ist es, objektive Daten zu generieren. Aber auch der Einfluss von nicht-teilnehmenden Personen auf das Beobachtungsgeschehen sollte sorgfältig abgeschätzt werden.
- *Offen vs. verdeckt*: Wissen die Beobachteten, z.B. die Lehrenden, dass sie beobachtet werden? Verdeckte Beobachtungen sind aus ethischen Gründen grundsätzlich bedenklich und im Zweifel sollte immer darüber informiert werden, wer, was und zu welchem Zweck beobachtet wird und was mit den erhobenen Daten anschließend geschieht.
- *Unstrukturiert vs. strukturiert*: Wie hoch ist der Grad an Standardisierung? Soll z.B. nur eine Strichliste geführt oder soll eine sprachliche Interaktion zwischen Teilnehmenden zusammenfassend notiert werden? Standardisierte Beobachtungen beugen Verzerrungen vor, fokussieren die Aufmerksamkeit der Beobachtenden und liefern zumeist quantifizierbare Daten, vernachlässigen jedoch wichtige Informationen jenseits des standardisierten Rasters.
- *Selbst vs. fremd*: Wird das eigene Verhalten im Sinne von „Introspektion“ oder das Verhalten von anderen beobachtet? Die Selbstbeobachtung unterliegt der Gefahr von starker Verzerrung und ist hochgradig störanfällig, ermöglicht jedoch einen guten Zugang zu Denkprozessen und Gefühlen.
- *Einzelne vs. Gruppen*: Wie viele beobachtende und wie viele beobachtete Personen gibt es? Je weniger Beobachtende beteiligt sind, desto weniger besteht die Gefahr der Beeinflussung der Geschehnisse, allerdings ist die Aufmerksamkeit von einzelnen Beobachtenden auf nur wenige Personen beschränkt.
- *Feld vs. Labor*: Findet die Beobachtung an einem natürlichen oder einem künstlich hergestellten Ort statt? Bei künstlichen Orten, z.B. in einem Seminar, ist nicht sicherzustellen, dass ein beobachtetes Verhalten auch außerhalb der künstlichen Situation, z.B. am Arbeitsplatz bestand hat.

Beobachtungen bieten die Möglichkeit, Daten über Prozesse und Gegebenheiten am Ort des Geschehens zu sammeln: „Durch Beobachtung erfasst man die Dinge dort,

wo sie passieren, in ihrem natürlichen Kontext, ungefiltert durch Berichterstatter oder Interviewpartner“ (Reischmann, 2006, S. 144). Beobachtungen bieten sich deshalb insbesondere an, wenn die Evaluation Informationen über das Seminargeschehen liefern soll, insbesondere über Interaktionsprozesse von Lehrenden und Lernenden oder auch von Lernenden untereinander. Nützlich werden sie nach Gerl (1983, S. 50), wenn den Lehrenden und interessierten Lernenden detaillierte Rückmeldungen über Formen und Inhalte der Interaktion gegeben werden können. Konkret bietet sich dafür seiner Meinung nach die Beobachtung des Handelns der Lehrenden in didaktischen Schlüsselsituationen an, wie etwa die gegenseitige Vorstellung gestaltet wird und welche Begründungen oder Zielsetzungen für den Einstieg in ein Teilthema genannt werden (vgl. ebd., S. 50-53). Um den Lernerfolg mithilfe von Beobachtungen zu ermitteln, lassen sich die Teilnehmenden während eines Seminars bei der Durchführung eines Rollenspiels oder anderer Übungen beobachten. Im betrieblichen Kontext kann man auch Verhaltensänderungen am Arbeitsplatz erheben (vgl. Häring, 2003, S. 87).

Nicht nur Personen, sondern auch Räumlichkeiten lassen sich beobachten. Reischmann (2006, S. 145-146) beschreibt ein Verfahren anhand eines Beispiels, bei dem eine externe Person die Zugänglichkeit von Räumen einer Bildungseinrichtung erkundet, indem sie den typischen Weg eines Teilnehmenden, angefangen von der Parkplatzsuche bis in den Seminarraum, „nachgeht“ und dabei alle Beobachtungen notiert, die ihr auffallen. Im von Reischmann geschilderten Beispiel kam kein standardisiertes Beobachtungsraster zum Einsatz, doch wurde versucht, zwischen einerseits Beobachtungen, die zum größten Teil introspektiv sind, und andererseits Kommentaren zu den Beobachtungen, eingeleitet durch das Kürzel „KB:“, zu unterscheiden: „Finde mich in einem dunklen Gang, weiß nicht ob dies richtiger Eingang, verlasse das Haus wieder ... es riecht nach Schule. KB: Erinnert mich an Schulstress, Noten-Angst. Warum wird hier nicht gelüftet?“ (ebd., S. 145). Ein derartiges Verfahren ist auch mit einem stärker standardisierten und mit Bewertungsskalen versehenen Beobachtungsraster vorstellbar.

Deutlich geworden ist, dass Beobachtung generell mit Aufwand verbunden ist, der zwar nicht bezüglich Planung, wohl aber hinsichtlich Durchführung deutlich über dem von schriftlichen Befragungen liegt. Systematische Beobachtungen stellen eine herausfordernde Aufgabe für die Beobachtenden dar. Diese müsse die Kompetenz besitzen, aus der Masse an beobachtbaren Informationen die relevanten herauszufiltern und dürfen diese weder durch eigene Interpretationen noch Selektionen verfä-

schen (vgl. Reischmann, 2006, S. 144-145; Beywl et al., 2007, S. 33; Diekmann, 2007, S. 551). Dies führt zu der Frage, wer eigentlich beobachtet. Ein Praktikant oder ein Kollege/eine Kollegin aus dem pädagogischen Team, wie von Gerl (1983, S. 48) genannt, sind sicherlich in der Lage, Beobachtungen durchzuführen und wertvolle Hinweise und Anregungen für das eigene Lehrverhalten zu geben, doch sind hierfür gegebenenfalls Schulungen notwendig.

### **2.5.3 Verwendung bereits vorliegender Materialien**

Während Befragungen und Beobachtungen Daten für die Evaluation produzieren, ist es auch möglich, auf bereits vorhandene Daten zurückzugreifen. Diese liegen in einer Weiterbildungsorganisation in großer Zahl und unterschiedlichsten Formen vor: Programmhefte, Webseiten, Berichte, Konzepte, Anträge, Entwürfe, Abrechnungen, Teilnehmendenlisten, Protokolle, Einladungen und Klassenbücher (vgl. Reischmann, 2006, S. 151-152). Hinzu kommen insbesondere auch Dokumente, die über erreichte Abschlüsse und Prüfungsergebnisse informieren wie z.B. Zeugnisse, aus denen sich Abschlussquoten, Notenspiegel und andere Informationen gewinnen lassen.

Die Vorteile der Verwendung bereits vorliegender Materialien für die Evaluation in der Weiterbildung liegen auf der Hand (vgl. u.a. Reischmann, 2006, S. 154): Die Daten stehen ohne aufwändige Datenerhebung, meist als Nebenprodukte zur Verfügung; sie ändern sich nicht, stellen authentische Quellen dar und sind durch die Erhebung nicht verfälscht worden. Darüber hinaus besteht vereinzelt sogar die Möglichkeit, Zeitreihen aus vorhandenen Daten zu erstellen. Interessant ist die Verwendung bereits vorliegender Materialien unter anderem auch deswegen, weil die Daten non-reaktiv, also ohne Beeinflussung des Gegenstands, gewonnen wurden. Nachteilig kann sich auswirken, dass die Daten eben nicht speziell für Evaluationszwecke erstellt worden sind und immer wohl überlegt werden muss, inwieweit die Daten tatsächlich geeignet sind, als Indikatoren und Kriterien bei der Beantwortung der gestellten Evaluationsfragen zu dienen (Stichwort: Validität).

### **2.5.4 Tests und Lernkontrollen**

Auch die Ergebnisse von Tests und Leistungskontrollen können in eine Evaluation einfließen. Ein Test dient dazu, den angestrebten Lernerfolg zu überprüfen, indem die Leistungen von Teilnehmenden mithilfe von Testinstrumenten bewertet werden, die hinsichtlich Aufgabenstellung, Durchführung und Auswertung klare Vorgaben beinhalten. Es handelt sich also um eine Form „objektivierter Leistungsmessung“

(Tietgens, 1971), denn im Hintergrund ist eine normative Vorstellung darüber vorhanden, was richtig und was falsch ist und der Test soll Informationen darüber liefern, ob bzw. inwieweit Teilnehmende dieser Vorstellung entsprechen. Damit wird auch das Einsatzgebiet von Tests deutlich: „Tests wendet man dort an, wo fest umschreibbares Wissen oder Können erworben werden soll“ (Reischmann, 2006, S. 150). Voraussetzung für die Durchführung eines Tests ist dementsprechend, dass klar definierte Lernziele vorliegen, die abgeprüft werden können (vgl. Bronner & Schröder, 1983, S. 208).

Tests können sich auf Wissen, auf Fertigkeiten oder auf beides beziehen, wobei die Überprüfung des Wissens meist schriftlich erfolgt, beispielsweise bei der Theorieprüfung der DLRG, um einen Schwimmschein zu erlangen oder bei einer Abschlussklausur für eine IT-Weiterbildungsmaßnahme. Tests zur Beurteilung erlangter Fertigkeiten durch eine Fort- und Weiterbildung finden in Form von praktischen Prüfungen statt, so etwa die Schwimmprüfung bei der DLRG oder die Abschlussprüfung bei einer Weiterbildung im handwerklichen Bereich. Der fließende Übergang zu den vorgenannten Erhebungsmethoden tritt dabei zu Tage: Schriftliche Prüfungen lassen sich in gewissem Sinne der Analyse vorhandener Materialien und Prüfungen von Fertigkeiten den Beobachtungsverfahren zuordnen.

Reischmann (2006, S. 148-149) unterscheidet zwischen standardisierten und informellen Tests. Standardisierte Tests sind meist schriftlich, haben einen langen Entwicklungsprozess durchlaufen, halten vorgegebenen Gütekriterien stand und ihre Messgenauigkeit, Mittelwerte und Streuungen sind für Vergleichszwecke bekannt. Sie liegen allerdings nur für eine geringe Themenanzahl vor, wenngleich sie für bestimmte Bereiche, wie die Berufseignung oder Berufsabschlussprüfungen unabdingbar sind. Zu einem informellen Test, der in der Weiterbildung häufiger anzutreffen ist, gehört zwar auch eine klare Aufgabenformulierung in Kombination mit einer verbindlichen Durchführung und Auswertung, der lange Entwicklungsprozess mithilfe statistischer Verfahren wird jedoch nicht durchlaufen. Die Ergebnisse werden nur im Nachhinein statistisch analysiert, wobei Aufgaben gegebenenfalls aus der Wertung entfernt werden.

Der Begriff „Test“ weckt schnell negative Assoziationen und ist mit zahlreichen Vorbehalten verbunden, da Tests unter anderem auf Selektion abzielen und Ängste erzeugen können (vgl. Bronner & Schröder, 1983, S. 208). Auch die Begriffe „Lernkontrolle“, „Lernüberprüfung“ oder „Performanz-Prüfung“ (Reischmann, 2006, S. 147) können einen negativen Beigeschmack haben. Buddenberg, Holzapfel, Uemming-

haus, Kaiser, Hohenstein und Wolter (2007, S. 48) bevorzugen daher den Begriff „Lernstandortbestimmung“, den sie insbesondere gegenüber Lernkontrolle abgrenzen. Eine von den genannten Autoren/innen durchgeführte Studie hatte zum Ergebnis, dass Teilnehmende der untersuchten Seminare grundsätzlich eine positive Tendenz gegenüber Lernstandortbestimmungen haben, sich allerdings Teilnehmende aus Kursen ohne beruflichen Verwertungszweck bei den Standortbestimmungen eher unwohl fühlten. Lernstandortbestimmungen wurde darüber hinaus eine lernunterstützende Funktion zugeschrieben, sofern ihr Einsatz im Lernsetting „kein methodischer Fremdkörper“ (Buddenberg et al., 2007, S. 49) sei. Insgesamt ermöglichten Lernstandortbestimmungen, Motivations- und Leistungspotenziale zu aktivieren, Reflexionsprozesse in Gang zu setzen und den Lerntransfer positiv zu beeinflussen.

Zu den allgemeinen Vorteilen von Tests zählen nach Reischmann (2006, S. 150-151) die Tatsache, dass objektive Fakten über das Wissen und Können vorliegen und dass sich die Qualität der Datenerfassung intersubjektiv überprüfen lässt. Wendet man den gleichen Test bei unterschiedlichen Gruppen an, können die Ergebnisse verglichen werden und schließlich unterstützt die Erstellung von Tests ihrerseits die didaktische Planung, da die Konstruktion klare Lernziele voraussetzt. Als Nachteile kommt zu den bereits genannten Ängsten, Vorbehalten und der geringen Verfügbarkeit von standardisierten Tests hinzu, dass die Konstruktion sehr aufwändig und zeitintensiv sein kann (vgl. Reischmann, 2006, S. 151; Häring, 2003, S. 84).

### **2.5.5 Auswahl und Bewertung unterschiedlicher Evaluationsmethoden**

Die Auswahl und Verwendung von Evaluationsmethoden wird durch die Vorstellungen darüber, was Evaluation ist und bezwecken soll, entscheidend beeinflusst. Will et al. (1987b) stellen zwei Prototypen von Evaluation gegenüber, in denen unterschiedliche Vorstellungen und Vorgehensweisen zum Ausdruck kommen: das Gutachtenmodell und das Helfer- und Beratungsmodell. Das *Gutachtenmodell* geht von Evaluationen aus, die wissenschaftlich orientiert sind, methodisch aufwändige Verfahren einsetzen, an einem möglichst hohen Grad an Objektivität, Reliabilität und Validität interessiert sind, summativ die Ziele und Wirkungen fokussiert und von externen, neutralen Evaluierenden durchgeführt wird. Als primäre Funktion steht dabei die Entscheidung im Vordergrund. Das *Helfer- und Beratungsmodell* versteht hingegen Evaluation als sozialen Interaktions- und Beratungsprozess, verzichtet auf die „Fiktion nonreaktiver Messung“ und sieht Evaluation stattdessen als reaktiven



Austauschprozess. Es ist auf kleinere Projekte ausgelegt, intendiert die Beteiligung aller Stakeholder und möchte schnell, praxisnah und handlungsorientiert im Sinne der responsiven Evaluation (vgl. Stake, 2004) sein. Hier steht als primäre Funktion die Steuerung und Optimierung im Vordergrund.

Von diesen unterschiedlichen Grundpositionen ausgehend, ergeben sich jeweilige Implikationen für die Auswahl der Evaluationsmethoden, die zum Einsatz kommen. Evaluationen in der Tradition des *Gutachtenmodells* sind empirisch-experimentell ausgerichtet und favorisieren Kontrollgruppendesigns. Statistische Verfahren stehen klar im Vordergrund, da ein Vergleich mit Normwerten und Standards beabsichtigt ist. Als Vorteile des Modells und der verwendeten Methoden sind die Möglichkeit der Quantifizierung und die Generalisierbarkeit zu nennen. Es ist also gut geeignet für umfangreiche und teure Programme, für die Effektivitätskriterien vorliegen, die mit Entscheidungen auf hoher Ebene verbunden sind und für die mathematisch-statistische Erkenntnisse benötigt werden. Dem gegenüber stehen als Nachteile, dass die aufwändigen Messverfahren nicht in der Lage sind, den Gegenstand vollständig und adäquat zu erfassen, und dass das methodische Procedere träge ist und schlecht auf Änderungen des evaluierten Gegenstands reagieren kann. Evaluationen, die dem *Helfer- und Beratungsmodell* folgen, setzen laut Will et al. auf ganzheitliche, qualitative Erfassungsmethoden wie narrative Interviews, teilnehmende Beobachtung, einfache Fragebögen auch mit offenen Antworten, Blitzlicht, Stimmungsbarometer, Wandzeitungen, Kurstagebuch, Gruppendiskussionen, Manöverkritiken. Auch Nebeneffekte sind von Bedeutung. Zu den Vorteilen gehören die Flexibilität der Methoden und Verfahren, die für komplexe, unstrukturierte Situationen geeignet sind. Die Ergebnisse haben hohe praktische Relevanz und die Akzeptanz unter den Beteiligten ist sehr hoch. Allerdings sind Evaluationen dieses Typs schwierig zu planen und die Ergebnisse sind nicht immer reproduzierbar und nicht generalisierbar. Daher eignet sich ein derartiges Vorgehen, wenn die Betroffenen in Entscheidungen einbezogen werden sollen, wenn ein hoher Grad an Objektivität nicht benötigt wird und wenn Programme noch nicht stark festgelegt sind. (Vgl. Will et al., 1987b, S. 30-34)

Während Will et al. es 1987 noch als Trend beschrieben, wie sich nach und nach der Blick von der outputorientierten Erfolgskontrolle hin auf eine Evaluation gerichtet hat, die alle Schritte des Bildungsprozesses berücksichtigt, hat sich in den letzten 25 Jahren das Bild weiter geändert. Die strikte Trennung der Prototypen ist zunehmend schwieriger herzustellen und es liegen Ansätze vor, die versuchen beide Positionen

stärker zu verbinden (wie z.B. bei Reischmann, 2006). Evaluationen, die eher dem Gutachtenmodell folgen, versuchen jetzt auch nicht-intendierte Wirkungen zu erfassen. Zudem kommen zunehmend Mixed-Methods-Methoden zum Einsatz – ein allgemeiner Trend, der in den Sozialwissenschaften zu beobachten ist und gelegentlich unter dem Begriff „third paradigm“ diskutiert wird (vgl. Kuckartz, 2009b). Dementsprechend formuliert Wessler (2010, S. 1043): „Es gibt weitgehend Konsens, dass qualitative und quantitative, non-formale und formalisierte Verfahren einander nicht ausschließen, sondern sich viel eher sinnvoll – und oftmals notwendigerweise – ergänzen.“ Und schließlich ist es nicht mehr nur noch politisch korrekt Beteiligte und Betroffene einzubeziehen, sondern dies hat sich als nutzenfördernd herausgestellt und wurde in Form von Standards für die Evaluation festgeschrieben, um die Nützlichkeit von Evaluationen zu sichern (vgl. DeGEval, 2008).<sup>32</sup>

Die bisherigen Ausführungen legen nahe, dass die Auswahl von Evaluationsmethoden stark durch das Grundverständnis von Evaluation und durch die verfolgten Evaluationszwecke vorbestimmt ist bzw. sein sollte. Wenn der Zweck einer Evaluation in der Bestimmung von Wirkungen liegt, fordert beispielsweise die DeGEval den Einsatz von (quasi-)experimentellen Designs mit Vergleichsgruppen, sofern es die Praxis erlaubt. Auch aus den Evaluationszeitpunkten, die eng mit den Zwecken verknüpft sind, lassen sich unterschiedliche Methodenpräferenzen ableiten: So bieten sich für eine Evaluation in der Planungsphase insbesondere die Inhaltsanalyse z.B. von Konzepten an, in der Durchführungsphase Befragungen und Beobachtungen, für die Ergebnisevaluation eher Befragungen und Tests, um den Lernerfolg zu messen.

Neben den Zwecken sind aber auch die einzelnen inhärenten Vor- und Nachteile der Methoden sowie die jeweils untersuchten Fragestellungen von ausschlaggebender Bedeutung, denn nicht jede Fragestellung lässt sich mit jeder Methode beantworten (vgl. Reischmann, 2006, S. 155). Häufig geht es deshalb um die Entscheidung für den Einsatz qualitativer, quantitativer oder Mixed-Methods-Methoden und deren jeweiligen Vorzüge. Gollwitzer und Jäger (2007) stellen hierzu (vornehmlich aus dem Blickwinkel einer traditionell quantitativen Forschungslogik) fest:

---

32 Guba und Lincoln (1989) arbeiten sehr detailliert heraus, wie Evaluation in den Jahren nach dem zweiten Weltkrieg vier Stufen durchlaufen hat, die jeweils im Vordergrund standen: (1) Kosten-Nutzen-Analyse, (2) Prozess-Orientierung, (3) Methoden-Orientierung und (4) Wirkungs-Orientierung. In den letzten Jahren hat sich eine fünfte Generation der Evaluation herausgebildet, die sich stärker an den Interessen der Beteiligten orientiert, um neben klaren Entscheidungen auch zu einem tieferen „Verständnis der komplexeren Qualitäten eines Programms“ (Wessler, 2010, S. 1032) zu kommen.

„Quantitative Methoden sind dabei für jene Fragestellungen eher geeignet, die hypothesentestend angelegt sind und bei denen eine Quantifizierung des Evaluationskriteriums inhaltlich sinnvoll erscheint (Wirksamkeit, Effizienz, ggf. Bedarf). Qualitative Methoden sind eher für jene Fragestellungen geeignet, die hypothesengenerierend angelegt sind und bei denen eine reine Quantifizierung dem Evaluationskriterium allein nicht gerecht würde (Problemanalyse, Konzeptionsqualität, Ausführungsintegrität usw.).“ (S. 151)

Gütekriterien stellen eine weitere adäquate Form zur Bewertung, und damit auch zur Auswahl, von Evaluationsmethoden dar. Zunächst müssen Methoden für die systematische Evaluation von Weiterbildungsprozessen entsprechend der in Abschnitt 2.2 vorgestellten Evaluationsdefinitionen dokumentierbar und auswertbar sein. Weiche Evaluationsverfahren wie Blitzlichter, Thermometerabfragen und vergleichbare Methoden können zwar den Lehr-Lern-Prozess positiv beeinflussen, eignen sich für eine Evaluation jedoch nur, wenn sie eine „Dokumentations- und Vergleichsmöglichkeit“ (Zech, 2008, S. 94) bieten. Diese Dokumentations- und Vergleichsmöglichkeit sei ist laut Zech für ein systematisches Qualitätsmanagement unverzichtbar, da nur so begründete und nachvollziehbare Konsequenzen für Entwicklungsmaßnahmen gezogen werden können (vgl. ebd., S. 94). Für die bei einer Evaluation eingesetzten quantitativen Methoden lassen sich die üblichen Gütekriterien Objektivität, Reliabilität und Validität, für Tests insbesondere die Normierungsmöglichkeit und die Ökonomie (Verhältnis von Aufwand zu Nutzen) heranziehen. Für die qualitativen Methoden eignen sich die von Mayring (2002, S. 145-149) vorgeschlagenen sechs Kriterien Regelgeleitetheit, Triangulation, Verfahrensdokumentation, kommunikative Validierung, argumentative Interpretationsabsicherung und Nähe zum Gegenstand.

Gerl (1983) verfolgt ein weniger systematisches und strenges Evaluationsverständnis, das stark auf die Reflexion im Lernprozess setzt, und gibt hierfür handlungsleitende Idealvorstellung zur Bewertung von Evaluationsmethoden an. Evaluationsmethoden sollen

1. die Selbstverantwortlichkeit der Teilnehmenden für das eigene Lernen bewusst machen,
2. den Unterschied zwischen reinem Beschreiben des Lernens und dem Bewerten des Lernens vor Augen führen,

3. allen Beteiligten Zeit zur Besinnung und Gelegenheit zur Klärung sowie Vergewärtigung ihrer Beobachtungen und Empfindungen geben,
4. allen Beteiligten dabei helfen, ihre Wahrnehmungen und Empfindungen auch ausdrücken zu können,
5. alle Beteiligten dabei unterstützen, Ausdrucksmöglichkeiten auf einer (möglichst) breiten Skala von Medien zu entwickeln,
6. dazu anregen, mögliche Lerngewinne auch in anderen Feldern als dem Bereich des thematisierten Lerngegenstandes zu bedenken und auszusprechen. Ebenso sollen, wenn es um die Evaluation von Lernprozessen geht, alle Wahrnehmungen oder Empfindungen in der gegenwärtigen Situation ausgedrückt werden können, also auch diejenigen, die nicht unmittelbar mit der Arbeit am Thema zu tun haben und
7. möglichst alle Beteiligten (auch die Lehrenden) aktiv in das Geschehen einbeziehen. (Vgl. Gerl, 1983, S. 25-28)

Diese Auflistung ist aufgrund seiner Prämissen nur bedingt auf alle Evaluationsmethoden anwendbar, jedoch besonders hilfreich für Evaluationen, die direkt den Lernprozess unterstützen sollen. Etwas allgemeiner kann man in diesem Zusammenhang nach Wesseler (2010, S. 43) das Gütekriterium anlegen, dass Evaluationsverfahren so ausgewählt werden sollten, „dass sie eine hohe Frequenz von Rückkopplung und Interaktion ermöglichen“.

Zusammenfassend kann gesagt werden, dass sich die Entscheidung für eine Evaluationsmethode von Weiterbildungsprozessen von einem theoretischen Standpunkt aus betrachtet (a) nach den verfolgten Zwecken, (b) nach den verfolgten Fragestellungen und (c) nach geeigneten Gütekriterien richten sollte. Allerdings ist zu vermuten, dass in der Praxis die Wahl stark durch andere Faktoren überlagert wird, insbesondere durch eher öffentliche wie die vorhandenen Ressourcen und das Know-how der Durchführenden und eher subtile wie Einrichtungskultur und Tradition. Dabei ist vermutlich davon auszugehen, dass die Wahl aufgrund pragmatischer Kriterien eine höhere Anschlussfähigkeit in einer Organisation hat als die Wahl aufgrund theoretischer Gesichtspunkte. Die DeGEval (2011) sieht in diesem Aspekt auch einen Unterschied zur Grundlagenforschung: „Anders als die Grundlagenforschung muss die meist auftragsbasierte Evaluation klug ausbalancieren zwischen hohen methodischen Standards und einer ökonomisch vertretbaren, pragmatischen, häufig unter Zeitdruck stehenden methodischen Herangehensweise“ (S. 2).

Folgende Fragen fassen abschließend die hier zusammengetragenen Kriterien und Hinweise zur Auswahl und Bewertung von Methoden zusammen:

- Sind die Methoden geeignet, um zur Beantwortung der Fragestellungen beizutragen und den Zweck der Evaluation zu erfüllen?
- Welchen Gütekriterien muss (bzw. kann überhaupt) die Methode genügen?
- „Welchem didaktischen Zwecke nutzen die erhobenen Daten bzw. das angewandte Evaluationsverfahren?“ (Reischmann, 2006, S. 160)
- Wie systematisch ist die Methode? Wie gut lassen sich ihre Ergebnisse dokumentieren? Inwieweit erlauben ihre Ergebnisse Vergleiche mit anderen Ergebnissen?
- Inwieweit sind reaktive Erhebungsmethoden vertretbar oder gar erwünscht?
- Inwieweit greift die Methode in den Lehr-Lern-Prozess ein; inwieweit fördert oder behindert und stört sie diesen? Inwieweit fördert sie das Lernen der Teilnehmenden?
- Wie viel verlangt die Methode den Datengebern/innen ab?
- Welche Ressourcen für Datenerhebung und Auswertung sind vorhanden?

## **2.6 Evaluationsansätze und -modelle mit Relevanz für die Weiterbildung**

Während in den bisherigen Ausführungen einzelne Aspekte von Evaluationen näher beleuchtet wurden, stehen in diesem Abschnitt Evaluationsansätze und -modelle im Vordergrund, welche sich nicht nur auf einzelne Ausschnitte, sondern auf den Evaluationsprozess als Ganzes beziehen und versuchen, diesem einen Orientierungsrahmen zu geben. Es liegt bis heute eine Fülle<sup>33</sup> unterschiedlicher Ansätze und Modelle der Evaluation vor – und wie Stockmann (2007, S. 40) bemerkt, beinahe genau so viele Versuche, diese Fülle zu systematisieren. Bevor ein Blick auf Systematisierungsversuche im Bereich der Weiterbildung geworfen wird, soll zuvor geklärt werden, was genau unter einem Evaluationsmodell zu verstehen ist. Beywl und Niestroj (2009, S. 40) definieren ein Evaluationsmodell als ein „ausformuliertes, theoretisch begründetes und durch praktische Evaluationserfahrungen gesättigtes Orientierungssystem, wie Evaluationen geplant und durchgeführt werden sollen“. Die Ansätze und Modelle stellen also „mehr oder minder ausformulierte Vorschläge zur prak-

---

33 Stufflebeam und Shinkfield (2007) unterscheiden beispielsweise 26 Ansätze.

tischen Strukturierung von Evaluationsvorhaben dar“ (Stockmann, 2007, S. 40). Evaluationsmodelle liegen daher auf einer höheren begrifflichen Stufe als die in diese Arbeit bisher getroffenen Unterscheidungen verschiedener Evaluationsaspekte, wie z.B. dem Zeitpunkt oder der Funktion von Evaluation (vgl. Wulf, 1972, S. 24). Es lässt sich zwar keine eindeutige Unterscheidung zwischen den beiden Termini „Ansatz“ und „Modell“ in der Literatur ausmachen, doch scheint der Begriff Ansatz von seiner Bedeutung weicher gefasst zu sein, der nicht so hohe Ansprüche an die Ausführungen stellt.

Ansätze und Modelle der Evaluation lassen sich unter anderem hinsichtlich

- ihrer Tragweite und möglicher Anwendungsfelder,
- ihrem Ausmaß an praktischer Handlungsanleitung,
- ihrer theoretischen Fundierung,
- ihren Aussagen über (zu verwendende) Methoden,
- ihrem Nutzenfokus und
- ihren Aussagen über den Einbezug der Beteiligten unterscheiden.

Wie bereits ausgeführt bestehen zahlreiche Versuche, die Ansätze und Modelle zu systematisieren, wie etwa bei Stockmann und Meyer (2010) oder Stufflebeam und Shinkfield (2007). Die verschiedenen Übersichten, die im Kontext der Weiterbildungsevaluation entstanden sind, unterscheiden sich hinsichtlich des primären Systematisierungsversuches. Während Frankenreiter (1996) fünf Modelle hinsichtlich ihrer Anzahl an Phasen bzw. Stufen unterscheidet, legt Götz (2001) eine nach Erscheinungsjahr chronologische Auflistung von fünf Ansätzen und Modellen und ihren jeweiligen Grundgedanken vor und Häring (2003) unterscheidet ganz allgemein pädagogische versus ökonomische Modelle. Unabhängig vom Systematisierungsverfahren ist auffällig, dass die meisten Ansätze und Modelle mit einer Person assoziiert sind (vgl. hierzu auch Stockmann, 2007, S. 41).

In den folgenden Unterkapiteln werden vier Ansätze bzw. Modelle vorgestellt, welche entweder einen direkten Bezug zur Weiterbildung und zu Weiterbildungsprozessen aufweisen, sich im Bereich Weiterbildung anwenden lassen oder die sich als allgemeines Ablaufmodell von Evaluationen gut für die später geplante Analyse eignen.

### 2.6.1 Der Vier-Ebenen-Ansatz von Kirkpatrick

Der Vier-Ebenen-Ansatz geht auf Donald L. Kirkpatrick zurück, der die grundlegenden Ideen zu diesem Ansatz in seiner Doktorarbeit mit dem Titel „Evaluating a Human Relations Training Program for Supervisors“ niedergeschrieben und in vier Artikeln in der Zeitschrift „Training und Evaluation“ im Jahr 1959 veröffentlicht hat. Auch wenn der Ansatz in den 60 Jahren seines Bestehens nur minimal von Kirkpatrick angepasst und erweitert wurde und nur eine geringe theoretische Fundierung aufweist, erfreut er sich bis heute einer großen Beliebtheit in Forschung und Praxis. Der Ansatz diente gleichermaßen unzähligen Forschungsstudien als Orientierungsrahmen (z.B. Häring, 2003) und zahlreichen Praktiker/innen als Praxisleitfaden. Im Jahre 2006 veröffentlichte Kirkpatrick gemeinsam mit seinem Sohn James die dritte Auflage seines Einführungswerkes „Evaluating Training Programs. The four levels“, das auf ca. 100 Seiten den Ansatz in einfacher und klarer Sprache darstellt und neuere Entwicklungen des Weiterbildungsfeldes wie Balanced Scorecard und E-Learning integriert. Den größten Teil des Buches nehmen auf ca. 250 Seiten 16 Fallstudien ein, die allesamt konkrete Umsetzungen des Vier-Ebenen-Ansatzes schildern. Wahrscheinlich ist es gerade diese Anwendungsorientierung, die sich ebenfalls durch die theoretischen Ausführungen zieht, – gepaart mit der Schlichtheit des Ansatzes – welche den vier Ebenen zu so großer Popularität verholfen hat.

Der gesamte Ansatz fokussiert stark auf den Kontext beruflicher Trainings, denen ein materieller Verwendungszweck innewohnt, weniger auf den Bereich von Bildung und Erziehung, der nicht primär einem Verwertungszweck unterliegt (vgl. Beywl, 2009, S. 127). Eine eindeutige Evaluationsdefinition findet sich im Standardwerk nicht, doch sieht Kirkpatrick seinen gesamten Ansatz selbst als Konkretion und Klärung verschiedener Evaluationsperspektiven und er liefert eine Beschreibung von Evaluation als Wissenschaft und Kunst: „As a science, it is organized knowledge – concepts, theory, principles, and techniques. As an art, it is the application of the organized knowledge to realities in a situation, usually with blend or compromise, to obtain desired practical results“ (Kirkpatrick & Kirkpatrick, 2006, S. 74). Zudem lokalisiert Kirkpatrick die Evaluation eines Programms am Ende eines zehnschrittigen Ablaufs, der von der Bedarfserschließung über die Auswahl geeigneter Lehrender bis hin zur Koordination des Programms führt. Kirkpatrick benennt drei klassische Evaluationszwecke: erstens die Optimierung zukünftiger Programme, zweitens die Schaffung von Entscheidungsgrundlagen, ob man ein Programm einstellt oder weiterführt, sowie drittens die Rechtfertigung von Investitionen in die Trainingsabtei-

lung eines Unternehmens (vgl. ebd., S. 19). Kern des Ansatz von Kirkpatrick bildet die Unterscheidung der folgenden vier Evaluationsebenen (vgl. Kirkpatrick & Kirkpatrick, 2006, S. 22-70; Einleitungssätze in Anlehnung an Beywl, 2009, S. 127):

1. *Die Reaktion (reaction)* der Teilnehmenden auf das Angebot als Ausmaß der Kundenzufriedenheit: Kirkpatrick betont die Wichtigkeit positiver Reaktionen auf ein Training, denn zum einen hänge hiervon die weitere Zukunft des Trainings ab und zum anderen reduziere eine negative Reaktion die Lernmotivation und die Chance auf weiteres Lernen. Als Methode kommen auf dieser Ebene vorwiegend kurze Fragebögen, sogenannte „happyness sheets“, zum Einsatz, die sowohl standardisierte Rückmeldungen über die Lehrenden, die Relevanz des Themas, die Trainingsgestaltung etc. als auch freie Kommentare und Vorschläge der Teilnehmenden erheben. Die Sheets dienen dazu, wertvolles Feedback zu erhalten und gleichzeitig den Teilnehmenden zu zeigen, dass man an ihren Einschätzungen und Bedürfnissen interessiert ist. Auch die Eindrücke der Koordinierenden sollten auf dieser ersten Ebene berücksichtigt werden.
2. *Das Lernen (learning)* der Teilnehmenden im Lernfeld, das heißt Zuwachs an Wissen, Veränderung von Einstellungen und Erwerb von Fertigkeiten gemäß der Programmziele: Lernen ist für Kirkpatrick die Voraussetzung für Verhaltensänderung und sollte ebenfalls evaluiert werden. Dies ist jedoch zeitaufwändiger und schwieriger als lediglich Reaktionen zu erheben. Pre-Post-Messungen dienen dazu, Veränderungen zu erheben, wobei sich Wissenszuwächse per Test, Einstellungsänderungen per Fragebogen (z.B. mithilfe standardisierter psychometrischer Managementtests) und die Verbesserung von Fertigkeiten mittels „performance tests“ erheben lassen. Sofern realisierbar, sollte eine Kontrollgruppe verwendet werden, um überzeugende Beweise vorlegen zu können.
3. *Das Verhalten (behavior)* der Teilnehmenden im Funktionsfeld, also Anwendung des Gelernten bei der Arbeit: Auf dieser Ebene steht das veränderte Arbeitsverhalten der Teilnehmenden im Fokus der Evaluation, das noch schwieriger zu erheben ist als Reaktionen und Lernen. Denn Voraussetzung für verändertes Arbeitsverhalten ist das Vorhandensein der Möglichkeit, dieses überhaupt zeigen zu können und wenn eine Möglichkeit vorliegt, lässt sich nicht vorhersagen, ob die Teilnehmenden das neue Verhalten zeigen. Es stellt sich somit die Frage nach Zeitpunkt, Häufigkeit und Art der Erhebung. Für Ebene drei bieten sich ebenfalls Pre-Post-Messungen sowie Kontrollgruppen an; alternativ können retrospektive Fragen nach Veränderungen weiterhelfen. Zudem sollten Personen interviewt



werden, die das Verhalten der Teilnehmenden einschätzen können, wie etwa deren Vorgesetzte und Mitarbeitende.

4. *Die Systemresultate (results)* im Unternehmen, z.B. bezüglich Mitarbeitendenfluktuation, Umsatz oder Gewinn: Die vierte Ebene ist zugleich die wichtigste und anspruchsvollste, da sie die Erreichung von Zielen adressiert, welche die Trainings letztendlich für das Unternehmen erbringen sollen. Wichtig ist die Ebene, weil es im Idealfall auf die Beantwortung der Frage herausläuft, welchen „Return on Investment“ ein Training abwirft, worunter vorwiegend Ergebnisse in Form belastbarer Zahlen wie Ausschussquoten und Krankenstände fallen, aber auch weichere Faktoren wie Jobzufriedenheit berücksichtigt werden können. Anspruchsvoll ist die Ebene, weil es methodisch schwer bis unmöglich ist, den Wirkungsnachweis zu führen. Die Messung der „results“ ist allerdings verhältnismäßig leicht zu realisieren, da zum Teil auf bereits vorliegende Unternehmensdaten zurückgegriffen werden kann.

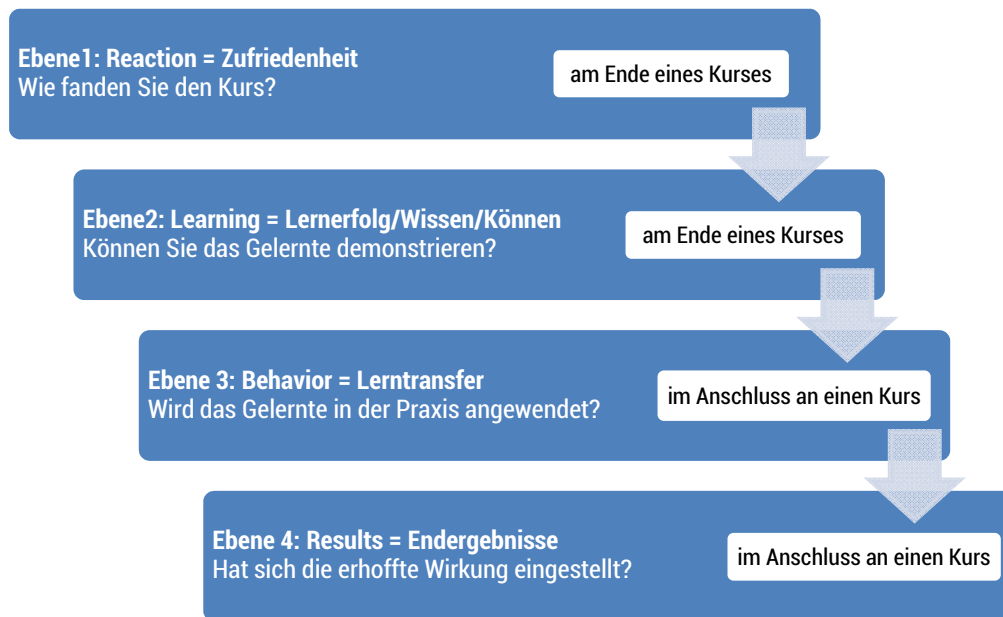
Sogenannte „Guidelines“ in Kirpatrick's Buch liefern konkrete Anleitungen und Hinweise für die Durchführung von Evaluationen auf dem jeweiligen Level. Auf allen Ebenen werden Standards festgelegt, anhand derer die Evaluationsergebnisse bewertet werden. Auf dem Reaction-Level kann man beispielsweise eine durchschnittliche Bewertung von 3,8 auf einer Skala von 0 bis 4 als kritischen Wert definieren, den Trainings überschreiten müssen. Wichtig zu betonen ist, dass bereits ab der zweiten Ebene elaborierte statistische und methodische Kenntnisse vonnöten sind, um die Evaluationen adäquat durchführen zu können.

Hervorzuheben ist darüber hinaus, dass die vier Ebenen als aufeinanderfolgende Stufen zu verstehen sind, wie Kirkpatrick und Kirkpatrick (2006) ausdrücklich betonen:

„The four levels represent a sequence of ways to evaluate programs. Each level is important and has an impact on the next level. As you move from one level to the next, the process becomes more difficult and time-consuming, but it also provides more valuable information.“ (S. 21)

Abb. 17 zeigt noch einmal die vier Ebenen – erweitert um den Zusammenhang zwischen Evaluationszeitpunkt und Evaluationsebene sowie ergänzt um weitere Übersetzungsmöglichkeiten der vier englischsprachigen Level-Bezeichnungen.

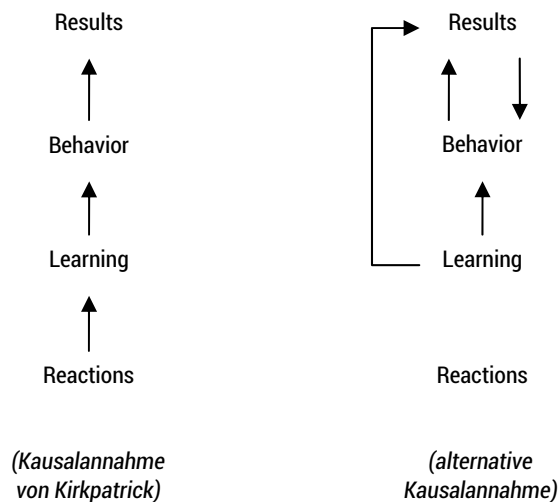
**Abb. 17: Die vier Evaluationsebenen von Kirkpatrick (eigene Darstellung, mit Ergänzungen nach Reichmann, 2006, S. 111-122; Balzer, 2005, S. 54-55)**



Trotz der großen Popularität sind auch einige kritische Aspekte zum Ansatz von Kirkpatrick aufzuzeigen. Am stärksten kritisiert und mehrfach untersucht wurde die von Kirkpatrick unterstellte Annahme, dass die vier Ebenen kausal aufeinander aufbauen und die jeweils untere Ebene einen Einfluss auf die nächste ausübt. Mehrere Studien bzw. Meta-Studien konnten nachweisen, dass zwischen der Zufriedenheit am Ende eines Kurses und dem Lernerfolg kein Zusammenhang bestehen muss (vgl. Alliger & Janak, 1989; Gessler & Sebe-Opfermann, 2011). Man kann auch erfolgreich lernen und dabei unzufrieden sein – und umgekehrt: zufriedene Teilnehmende müssen nicht zwangsläufig etwas gelernt haben. Geringe Korrelationen zeigen sich jedoch zwischen der zweiten und dritten, der zweiten und vierten sowie der dritten und vierten Ebene, was Alliger und Janak zu dem Entwurf eines Alternativmodells veranlasste, welches die Zufriedenheits-Reaktionen in der Kausalkette außen vorlässt und in Abb. 18 dargestellt ist.<sup>34</sup>

<sup>34</sup> Die Autoren weisen darauf hin, dass die erhobenen Daten keine endgültige Antwort auf die Frage erlauben, welches der Modelle einen korrekten kausalen Zusammenhänge darstellt. Im Angesicht der Ergebnisse sei ihr Vorschlag jedoch plausibel und könne zumindest zukünftige Forschung stimulieren (vgl. Alliger & Janak, 1989, S. 337-338).

**Abb. 18: Der kausale Zusammenhang zwischen den vier Kirkpatrick-Ebenen**  
(Quelle: Alliger & Janak, 1989, S. 335)



Mit Blick auf die zeitliche Perspektive in der weiter oben stehenden Abb. 17 wird deutlich, dass der Ansatz erst am Ende einer Bildungsmaßnahme greift und den Lehr-Lern-Prozess oder gar die Planung einer Veranstaltung unberücksichtigt lässt.<sup>35</sup> Evaluationen nach diesem Modell sind daher per se stärker summativ ausgerichtet, wenngleich die Zufriedenheitsabfragen auch zahlreiche Aspekte für Verbesserungen liefern können. Dieses Defizit hat Reischmann (2006, S. 111) zum Anlass genommen, den vier Ebenen eine erste Stufe noch vor der Kursauswahl voranzustellen, die er „Angebots-Evaluation“ und „Vorab-Evaluation“ nennt, und in der er fragt, ob und inwieweit ein Kurs das vorliegende Problem lösen wird.

Einen umgekehrten Weg beschreitet Phillips, der den vier Stufen von Kirkpatrick eine fünfte Stufe hinzufügt, die er mit „ROI – Return on Investment“ überschreibt und die als „aussagekräftigste Stufe der Evaluation ... den monetären Nutzen des Programms mit seinen Kosten“ (Philipps & Schirmer, 2005, S. 28) vergleicht. Die Evaluation wird hierdurch hochgradig ökonomisiert und das Evaluationsergebnis besteht am Ende aus einer Prozentzahl oder einer Nutzen/Kosten-Relation. Das Modell rückt den Unternehmensnutzen gänzlich in den Vordergrund, während der Nutzen für die sich bildenden Individuen in den Hintergrund rückt.<sup>36</sup>

<sup>35</sup> Diese Aspekte werden jedoch in den neun didaktischen Schritten berücksichtigt, die nach Kirkpatrick vor der Evaluation stattfinden.

<sup>36</sup> Der Ansatz von Kirkpatrick und Philipps wurde von Kellner (2006) unter dem Titel „Value-of-Investment“ noch weiter ausgebaut. Das Modell von Kellner erweitert den Return on Investment um die qualitative Wertschöpfungskette von Weiterbildungsmaßnahmen und versucht im Kontext von Bildungscontrolling im betrieblichen Bereich eine ganzheitliche Messung der Weiterbildungsprozesse vorzunehmen.

### 2.6.2 Das CIPP-Modell von Stufflebeam

Das CIPP-Modell wurde von Daniel L. Stufflebeam Mitte der 1960er Jahre in den USA entwickelt. Die Wurzeln des Modells liegen in der Schulpädagogik: Mit Blick auf Evaluationen von Projekten für benachteiligte Schüler/innen an öffentlichen Schulen übte Stufflebeam scharfe Kritik an den damals üblichen Programmevaluationen, die mithilfe experimenteller Designs, standardisierter Tests und dem Fokus auf Zielerreichungsmessungen durchgeführt wurden (vgl. Stufflebeam, 1969; 1972 übersetzt ins Deutsche). Stufflebeam kritisierte, dass die Schulen mithilfe derartiger Evaluationen die angedachten Projekte nicht sinnvoll gestalten und nicht an den pädagogischen Erfordernissen der Kinder ausrichten könnten. Dies veranlasste ihn dazu, das CIPP-Modell zu entwerfen, das die Evaluation von innovativen und entwicklungsorientierten Bemühungen unterstützen und dabei pädagogische Entscheidungen berücksichtigen sollte (vgl. Stufflebeam, 2003, S. 30). CIPP ist ein Akronym für die vier Evaluationsbereiche Context, Input, Process und Product. Es lassen sich dementsprechend vier Evaluationstypen unterscheiden (vgl. Stufflebeam, Madaus & Kellaghan, 2000; Stufflebeam, 2003, 2007; Stufflebeam & Shinkfield, 2007):

*Context-Evaluationen* beantworten die Frage „What needs to be done?“ Sie bestimmen Bedarfe, Ressourcen und Probleme bezüglich eines Gegenstandes und helfen beispielsweise dabei, dessen Ziele zu definieren und zu priorisieren. Eine Bedarfserschließung in der Weiterbildung zur Ermittlung von Kundenbedarfen, wie sie auch das LQW-Modell vorsieht (vgl. Abschnitt 3.1), lässt sich beispielsweise als eine Context-Evaluation auffassen.

*Input-Evaluationen* beantworten die Frage „How should it be done?“ Sie beurteilen alternative Ansätze, konkurrierende Strategien, Ablaufpläne und Finanzen. Input-Evaluationen werden beispielsweise verwendet, um sich für eine Realisierungsart, eine Stellenbesetzung und für eine geeignete Geldzuweisung zu entscheiden. Die im vorigen Abschnitt erwähnte Vorab-Evaluation von Reischmann, die vor der Durchführung einer Bildungsmaßnahme durchgeführt wird, lässt sich als eine solche Input-Evaluation klassifizieren.

*Process-Evaluationen* beantworten die Frage „Is it being done?“ Sie erfassen, dokumentieren und bewerten die Implementation eines Programms. Sie helfen den Programmverantwortlichen bei der konkreten Umsetzung der geplanten Aktivitäten, indem sie Informationen darüber liefern, ob die Aktivitäten fristgerecht, planmäßig, finanzgemäß und effektiv ablaufen. Dadurch ermöglichen sie es, bei Bedarf die Umsetzung eines Programms anzupassen.

*Product-Evaluationen* beantworten die Frage: „Is it succeeding?“ Sie identifizieren und beurteilen die intendierten, nicht-intendierten, kurz- und langfristigen, positiven wie negativen Wirkungen eines Programms. Dadurch helfen sie den Erfolg bezüglich ausgewählter Bedarfe abzuschätzen und liefern Entscheidungsgrundlagen für die Anpassung, Weiterführung oder Beendigung eines Programms.<sup>37</sup>

Die vier Evaluationstypen nehmen unterschiedliche Zeitkorridore eines Programms in den Blick. Während Context- und Input-Evaluationen speziell auf die Zeit vor der Durchführung eines Projekts fokussieren, konzentriert sich die Process-Evaluation auf die Durchführung selbst und die Product-Evaluation blickt auf die Phase nach der Durchführung. Wenngleich unterschiedliche Zeitkorridore im Blick sind, heißt dies jedoch nicht, dass die Evaluationstypen zu einem bestimmten Zeitpunkt durchgeführt werden. Das CIPP-Modell versteht sich als „comprehensive framework“ (Stufflebeam, 2003, S. 2), das sowohl formative (prospektiv, begleitend), wie auch summative Evaluationen (retrospektiv, abschließend) anleiten kann. Für summative Evaluationen werden die bei den vier Typen der Evaluation gestellten Fragen in die Vergangenheit gesetzt: „Were important needs addressed (context)? Was the effort guided by a defensible plan and budget (input)? Was the service design executed competently and modified as needed (process)? Did the effort succeed (product)?“

Es ist wichtig zu erwähnen, dass nicht alle vier Evaluationstypen zum Einsatz kommen müssen. Vielmehr ist abzuwägen, welche der vier Typen adäquat in einer Situation sind. Drei Faktoren sind hierbei bedeutsam. Erstens, wann mit einer Evaluation begonnen wird: Wenn eine Bildungsmaßnahme bereits läuft, lässt sich zwar eine retrospektive Context- und Input-Evaluation durchführen, man wird sich aber aus pragmatischen Gründen eher auf Process- und Product-Evaluationen konzentrieren. Zweitens, welcher Evaluationszweck verfolgt wird: Wenn summative Evaluationen durchgeführt werden sollen, wird man eher alle vier Typen einsetzen, während bei formativen Evaluationen gelegentlich nur auf einzelne Aspekte zur Entscheidungsfindung fokussiert wird. Drittens, welche Menge an Informationen zu den einzelnen CIPP-Komponenten bereits vorliegt: Wenn zum Context bereits viele Informationen vorliegen, kann man eventuell auf tiefergehende Context-Evaluation verzichten. (Vgl. Stufflebeam, 2003, S. 19)

---

37 Bei Evaluationen, die Langzeiteffekte untersuchen, kann die Product-Komponente aufgegliedert werden in Wirkungen, Effektivität, Nachhaltigkeit und Übertragbarkeit (vgl. Stufflebeam, 2003).

Stufflebeam hat ein sehr umfassendes Evaluationsverständnis. Er definiert Evaluation als (a) einen Prozess des Erschließens, Einholens, Berichtens und Anwendens von beschreibenden und beurteilenden Informationen über (b) die Güte, die Tauglichkeit, die Redlichkeit und die Bedeutsamkeit eines Gegenstands, um (c) Entscheidungsprozesse anzuleiten, einen verantwortlichen Umgang zu unterstützen, effektive Strategien zu verbreiten und das Verständnis der beteiligten Phänomene zu erhöhen (vgl. Stufflebeam, 2003, S. 10). Zwar wurde das CIPP-Modell ursprünglich zur Unterstützung pädagogischer Entscheidungen entwickelt, doch hebt Stufflebeam nicht die Entscheidungsorientierung als Zweck von Evaluationen hervor, sondern positioniert das Modell klar unter den verbesserungsorientierten Modellen und betont: „The most important purpose of evaluation is not to prove, but to improve“ (Stufflebeam & Shinkfield, 2007, S. 331). Das CIPP-Modell fordert die Beteiligten zudem dazu auf, die Werte, die einem Gegenstand und seiner Evaluation zugrundeliegen, zu identifizieren und zu klären. Stufflebeam (2007, S. 333) benennt zahlreiche Beispiele für Werte im Schulbereich, etwa dass die speziellen Gaben und Talente eines Kindes gefördert oder dass zur Verfügung stehende Ressourcen effektiv verwendet werden sollen.<sup>38</sup> Aus den Werten lassen sich Evaluationskriterien ableiten, welche wiederum die Grundlage für die Auswahl der verwendeten Methoden darstellen. Bereits zu Entwicklungsbeginn wurde großer Wert auf die Berücksichtigung von Evaluationsstandards gelegt, die immer auch Nützlichkeitsaspekte umfassten (vgl. Stufflebeam, 1971).

Bezüglich der Methodenwahl steht im CIPP-Modell die gesamte Fülle qualitativer und quantitativer Methoden gleichberechtigt nebeneinander, wobei sich für unterschiedliche Evaluationstypen zum Teil dieselben, zum Teil verschiedene Methoden anbieten.<sup>39</sup> Der Begriff der Methodenpluralität erscheint hier angemessen, nennt Stufflebeam doch unter anderem Umfragen, Literaturrecherchen, Dokumentanalysen, Visitationen anderer Programme, Delphi-Techniken, Beobachtungen, Stakeholder-Interviews, Gruppendiskussionen und Fotoaufnahmen. (Vgl. Stufflebeam, 2003, S. 16)

---

38 Werte sind auch im LQW-System von Bedeutung und finden sich beispielsweise im Leitbild einer Organisation wieder oder spiegeln sich in der Definition gelungenen Lernens (vgl. Abschnitt 3.1).

39 Eine detaillierte tabellarische Übersicht einsetzbarer Methoden bei den einzelnen Evaluationstypen findet sich bei Stufflebeam (2003, S. 16).

Stufflebeam betont mehrfach die Praxisnähe des Modells und verweist auf zahlreiche Checklisten, die er im Internet bereitstellt, unter anderem eine sehr ausführliche zum CIPP-Modell selbst (zu finden unter [www.wmich.edu/evalctr/checklists](http://www.wmich.edu/evalctr/checklists)).<sup>40</sup>

Das CIPP-Modell wurde zwar Ende der 1960er Jahre ursprünglich im Kontext von Schulreformen entwickelt, hat jedoch seit seines 40-jährigen Bestehens weltweite Verbreitung gefunden und kommt in zahlreichen Feldern zur Anwendung. Das Modell ist sehr allgemein gehalten und lässt sich daher im Bereich der Weiterbildung sowohl auf der Makroebene des Weiterbildungssystems, auf der Mesoebene einer Weiterbildungsorganisation als auch auf der Mikroebene von Lehr-Lern- bzw. Bildungsprozessen anwenden. Will et al. (1987a, S. 19) haben das CIPP-Modell auf die berufliche Aus- und Weiterbildung übertragen, wobei sie jedem Schritt auf Ebene der Bildungsmaßnahme ein Evaluationsfeld zuordnen: Bedarfsermittlung, Zielerklärung > Kontext- und Ziel-Evaluation; Entwicklung, Vorbereitung > Input-Evaluation; Durchführung > Prozess-Evaluation; Ergebnisfeststellung im Lernumfeld und gegebenenfalls in der Anwendungsphase > Produkt-Evaluation. Taylor und Beniest (2003) formulieren zahlreiche konkrete Fragen, die auf der jeweiligen CIPP-Ebene gestellt werden können, und verdeutlichen dadurch die konkrete Übertragung und Anwendung des Modells auf eine Bildungsveranstaltung. Im Folgenden sind zu allen vier Ebenen einige der von Taylor und Beniest genannten Fragen wiedergegeben:

**„Context** What is the relation of the course to other courses? Is the time adequate? What are critical or important external factors (network, ministries)? Should courses be integrated or separate [sic!]? What are the links between the course and research/extension activities? Is there a need for the course? Is the course relevant to job needs?

**Inputs** What is the entering ability of students? What are the learning skills of students? What is the motivation of students? What are the living conditions of students? What is the students' existing knowledge? Are the aims suitable? Do the objectives derive from aims? ...

**Process** What is the workload of students? How well/actively do students participate? Are there any problems related to teaching? Are there any problems

---

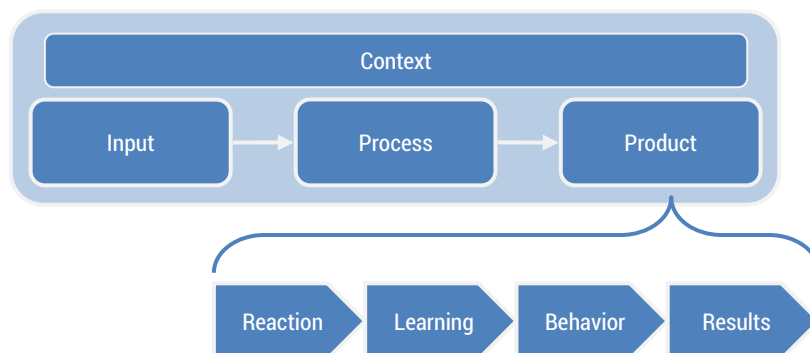
40 Die Checkliste zum CIPP-Modell wurde von Wolfgang Beywl ins Deutsche übersetzt und ist im Internet verfügbar: [www.univation.org/download/Checkliste\\_Stufflebeam\\_CIPP.pdf](http://www.univation.org/download/Checkliste_Stufflebeam_CIPP.pdf).

related to learning? Is there effective 2-way communication? Is knowledge only transferred to students, or do they use and apply it? ...

**Product** Is there one final exam at the end or several during the course? Is there any informal assessment? What is the quality of assessment (i.e. what levels of KSA [knowledge, skills, attitudes] are assessed?) What are the students' KSA levels after the course? ..." (S. 175-177; im Original als Aufzählung formatiert)

Im Vergleich zu dem auf betriebliche und berufliche Bildung fokussierten Ansatz von Kirkpatrick ist das CIPP-Modell aufgrund seiner breiten Anwendbarkeit weniger konkret. Der Vier-Ebenen-Ansatz lässt sich in das CIPP-Modell einbetten, wie Abb. 19 zeigt. Die vier Ebenen beziehen sich primär auf die Product-Komponente, die sich entsprechend untergliedern lässt.

**Abb. 19: Integration des Vier-Ebenen-Ansatzes in das CIPP-Modell (eigene Darstellung)**

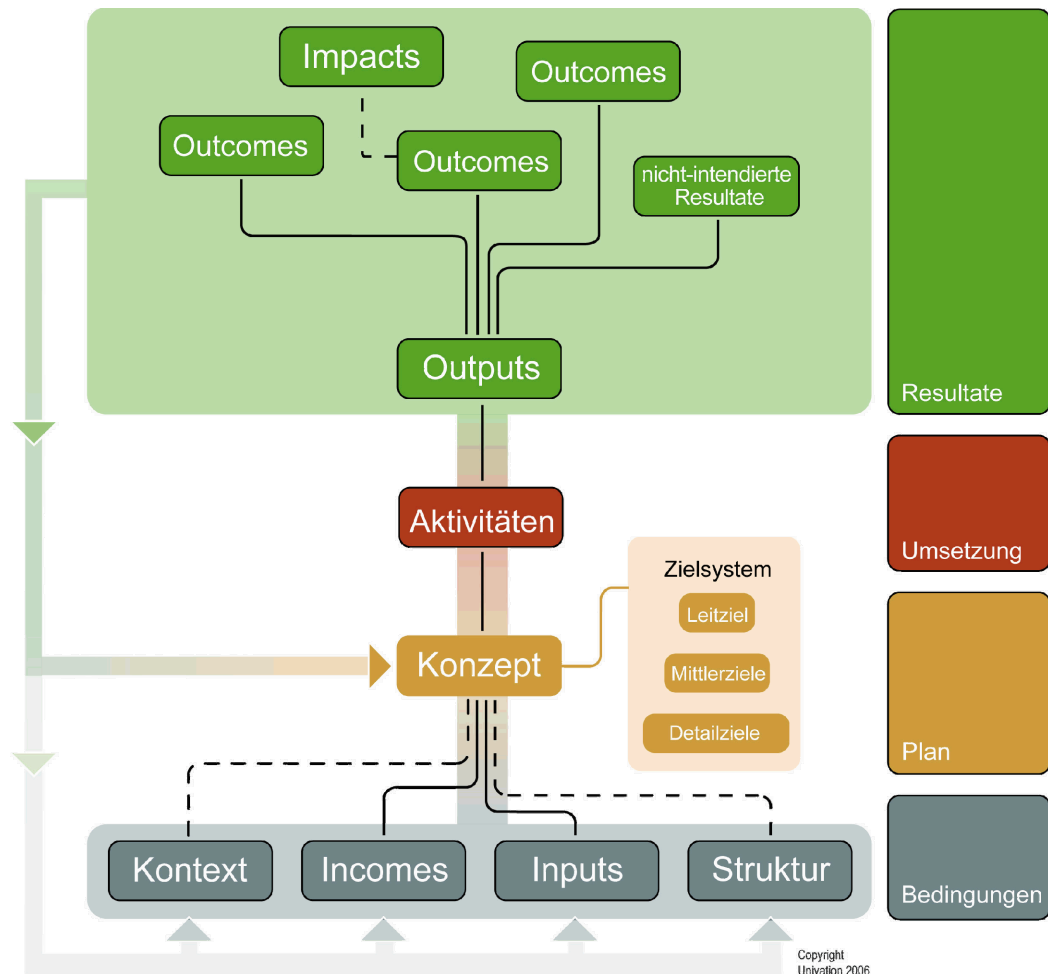


### 2.6.3 Der Programmbaum von Beywl et al.

Der Programmbaum ist ein sogenanntes „logisches Programmmodell“, das von Wolfgang Beywl und Univation, einem in Köln ansässigen privatwirtschaftlichen Evaluationsinstitut, entwickelt wurde. Abb. 20 stellt den Programmbaum visuell dar und verdeutlicht die Bedeutung eines logischen Modells in der Evaluation: Es dient der visuellen Veranschaulichung der Ablauflogik eines Programms, indem es angenommene Verbindungen (ohne zwangsläufige theoretische Begründung) einzelner Elemente verdeutlicht (vgl. Beywl & Niestroj, 2009, S. 67).



Abb. 20: Visuelle Darstellung des Programmbaums

(Quelle: [http://www.univation.org/download/Programmbaum\\_mit\\_Boxen.gif](http://www.univation.org/download/Programmbaum_mit_Boxen.gif))

Wie Abb. 20 zeigt, besteht der Programmbaum aus den vier Ebenen Bedingungen, Plan, Umsetzung und Resultate. Im Folgenden werden die einzelnen Bestandteile der vier Ebenen kurz beschrieben (verkürzte und leicht adaptierte Darstellung einer tabellarischen Übersicht einschließlich Beispiel von Beywl & Nistroj, 2009, S. 141-144). Zur untersten Ebene der *Bedingungen* zählen folgende Elemente:

- **Kontext** – Gesamtheit der umgebenden Systemumwelt eines Programms je nach dessen Reichweite auf lokaler, nationaler und/oder internationaler Ebene (z.B. Gesetze wie SGB II; Anteile von Arbeit suchenden Personen an der Bevölkerung einer Region).
- **Incomes** – Ressourcen, die Mitglieder der Zielgruppen in das Programm einbringen (z.B. Lerneingangsbedingungen der Teilnehmenden von Qualifizierungsmaßnahmen, wie Vorwissen, Motivation etc.; körperliche Verfassung).
- **Inputs** – Finanzielle, personale oder andere Ressourcen, die in ein Programm investiert werden, teils als monetäre Kosten erfassbar (z.B. Personal- und Sach-

aufwendungen, Anzahl und Qualifikationen der im Programm Tätigen, etwa spezielle Fortbildungen in lösungsorientierter Beratung).

- Struktur – Bedingungen, die beim Träger eines Programms vorliegen (z.B. Kooperationen, Rechtsform Kapitalausstattung, Leitbild).

Die zweite Ebene *Plan* umfasst folgende Elemente:

- Konzept – Gedanklicher (Grob-)Entwurf eines Programms dazu, mit Hilfe welcher Interventionen welche Resultate bis wann, wo bzw. bei wem ausgelöst werden sollen und wie der Programmprozess insgesamt gesteuert und überwacht werden soll (z.B. Förderantrag für eine Qualifizierungsmaßnahme oder ein Profiling-Angebot; Curricula für Weiterbildungsangebote).
- Zielsystem – Gesamtheit expliziter, schriftlich dargelegter Ziele für ein Programm auf den drei Konkretionsstufen Leitziele, Mittlerziele und Detailziele. Ein Leitziel ist ein weit gestecktes, abstraktes und über lange Zeit stabiles Ziel (z.B. die Beschäftigungsfähigkeit von Jugendlichen nach dem Hauptschulabschluss ist erhöht und die Lehrstellensuche ist erfolgreich). Mittlerziele verbinden Leitziele mit den Detailzielen und gliedern bzw. konkretisieren die Schwerpunkte eines Leitziels (z.B. Jugendliche nach dem Hauptschulabschluss meistern die für eine Bewerbung erforderlichen Schritte und verfügen über die erforderlichen Instrumente). Detailziele konkretisieren Mittlerziele, die im Idealfall spezifisch, messbar, akzeptabel, realistisch und terminiert (S.M.A.R.T.) sind (z.B. alle Jugendlichen der Klasse haben vor Erhalt des Abschlusszeugnisses der Hauptschule eine aktuelle, die individuellen Stärken klar benennende Bewerbungsmappe erstellt).

Auf der dritten Ebene ist die Umsetzung angesiedelt, die nur ein Element enthält:

- Aktivitäten – Handlungen der beruflich oder ehrenamtlich im Programm Tätigen: (a) Interventionen, die sich auf die Zielerreichung richten (z.B. Beratung, Profiling), (b) Hilfsaktivitäten, welche die Interventionen ermöglichen oder unterstützen (z.B. Teamaufbau, Einarbeiten neuer Mitarbeitenden), und (c) Zielgruppen-Aktivitäten, die eine aktive Beteiligung der Zielpersonen erfordern (z.B. Lernen, Bewerben, ...).

Die oberste Ebene der Resultate ist umfassend aufgegliedert und umfasst fünf Elemente:

- Outputs – In unmittelbarem Zusammenhang mit den Interventionen eines Programms stehende Resultate in Form (zählbarer) Leistungen/Produkte, Teilnahme-/Benutzungsmengen sowie Zufriedenheit der Benutzenden/Zielgruppen (z.B. Anzahl Teilnehmendenstunden, Anzahl Beratungsgespräche).
- Outcomes – Intendierte Resultate bei Zielgruppen, wie etwa Veränderungen bzw. Stabilisierungen im Wissen, den Einstellungen, sozialen Werten oder dem Können, im Verhalten oder in der Lebenslage/dem Status der Zielpersonen (z.B. Erweitertes Wissen über Bewerbungsmöglichkeiten, verstärkte Bewerbungstätigkeit, Integration in den Arbeitsmarkt).
- Impacts – Resultate eines Programms, die über die bei Zielgruppen auftretenden Outcomes hinausgehen, etwa in Form von Merkmalen sozialer Systeme, einer Organisation, eines Sozialraums, eines regionalen Weiterbildungssystems, der Bildungspolitik eines Bundeslandes (z.B. verbesserte Zusammenarbeit der Arbeitsverwaltung mit Weiterbildungsanbietenden).
- nicht-intendierte Resultate – Resultate eines Programms auf verschiedenen möglichen Ebenen, die außerhalb der im Konzept festgehaltenen Ziele liegen (z.B. steigt die Wohn-Attraktivität eines Stadtteils, in dem intensiv Maßnahmen zur Beschäftigungsförderung betrieben wurden, wodurch der Wohnungsmarkt und die Mietpreise beeinflusst werden).

Wie in der visuellen Darstellung des Programmbaums in Abb. 20 ersichtlich ist, greift der Programmbaum Elemente des CIPP-Modells nach Stufflebeam auf.<sup>41</sup> Die Context-Komponente von Stufflebeam wird dabei aufgegliedert in die vier Aspekte Kontext, Incomes, Inputs und Struktur. Der Process entspricht im Wesentlichen den Aktivitäten und die Products-Komponente der zeitlich und inhaltlich aufgeteilten Ebene der Resultate.

Beywl und Niestroj (2009, S. 140) sehen einen umfassenden Nutzen bei der Anwendung des Programmbaums. Konkretisiert man den von ihnen genannten Nutzen für die Weiterbildung, dann hilft dieser Baum-Ansatz den Verantwortlichen, Mitarbeitenden und Lehrenden zum einen dabei, die durchgeführten Bildungsveranstaltungen zu planen und zu steuern, indem sie die Logik der Bildungsveranstaltungen entschlüsseln und dabei die Voraussetzungen, Aktivitäten und Resultate in Beziehung zueinander setzen. Zum anderen kann der Programmbaum für die Kommunikation

---

41 Das CIPP-Modell stellt übrigens ebenfalls ein Programmmodell dar, da es Context, Input, Process und Product in Beziehung zueinander setzt.

über das Konzept der Bildungsveranstaltung eingesetzt werden, etwa um die verwendeten Lehrmethoden und die erwarteten Resultate bei der Zielgruppe anzupreisen. Evaluierende, die im Kontext der vorliegenden Arbeit in der Regel die für das Bildungsprogramm Verantwortlichen und die Lehrenden selbst sind, können durch die Anwendung des Programmbaums die Bildungsveranstaltung systematisch beschreiben, die generelle Evaluierbarkeit der Bildungsveranstaltung bzw. ausgewählter Aspekte abschätzen sowie gesammelte Daten, Informationen und Schlussfolgerungen den einzelnen Ebenen und Komponenten zuordnen.

Das zugrundeliegende Evaluationsverständnis hat Beywl zusammen mit Kollegen und Kolleginnen an mehreren Stellen ausgeführt, wobei Evaluation ausdrücklich als Dienstleistung aufgefasst wird:

„Evaluation als Dienstleistung für die Beteiligten eines Programms schafft datenbasiert Transparenz, sie präzisiert und dokumentiert, was innerhalb eines Programms passiert und welche Resultate entstehen. Die Evaluationsergebnisse sollen dem Programm nutzen, z.B. zur Weiterentwicklung der Ziele oder/und zur Optimierung der Aktivitäten.“ (Beywl et al., 2007, S. 6)

Der Programmbaum ermöglicht es im Besonderen wirkungsorientierte Evaluationen durchzuführen, die explizit und detailliert die Resultate in den Blick nehmen. Zu diesem Zweck sind im Zusammenhang mit dem Programmbaum weitere Hilfsmittel vorhanden, wie etwa die „Resultate-Treppe“, welche die Resultate eines Programms, z.B. einer Bildungsveranstaltung, in drei unterschiedliche Output-Stufen, zwei Outcome-Stufen und eine Impact-Stufe unterteilt (vgl. Beywl & Niestroj, 2009, S. 145).

## **2.6.4 Allgemeines Ablaufmodell von Evaluationen**

Die bisher vorgestellten Ansätze und Modelle lassen noch offen, wie eine Evaluation konkret abläuft. Insbesondere mit Blick auf die geplante spätere Auswertung bietet es sich daher an dieser Stelle an, ein allgemeines Ablaufmodell von Evaluationen vorzustellen, das sich auf möglichst viele Praxissituationen übertragen lässt. Vor diesem Hintergrund wird im Folgenden ein Ablaufmodell bestehend aus sechs Schritten beschrieben, welches sich unter anderem an der Arbeit von Balzer (2005) orientiert. Balzer hat empirisch die Frage untersucht, wie Evaluationsprojekte erfolgreich werden und aufbauend auf den Ergebnissen einen integrierenden theoretischen Ansatz vorgelegt. Ergänzend werden im Folgenden eigene frühere Ausführun-

gen (Rädiker, 2010c) sowie die Schritt-für-Schritt-Anleitung von Beywl et al. (2007) berücksichtigt.

### *1. Schritt: Evaluationsgegenstand, Gegenstandsziele und Stakeholder bestimmen*

Als erster Schritt wird festgelegt, welcher Gegenstand im Fokus der Evaluation stehen soll. Dieser Evaluationsgegenstand wird im Detail beschrieben und in der Regel gilt es dann, die Ziele, die der Gegenstand verfolgt bzw. die mit ihm erreicht werden sollen, zu identifizieren. Um die Werthaltungen, Bedarfe und Positionen relevanter Stakeholder bei der Evaluation berücksichtigen zu können, sollte zudem eruiert werden, welche Personen(gruppen) am Gegenstand beteiligt und welche von den Evaluationsergebnissen betroffen sind.

### *2. Schritt: Evaluationszweck, Fragestellungen und Bewertungskriterien bestimmen*

Im zweiten Schritt gilt es, den Evaluationszweck festzulegen, der als roter Faden zur Orientierung und Ausrichtung der Evaluation dienen soll. Dabei gilt es zu beachten, dass verschiedene Stakeholder unterschiedliche Zwecke mit einer Evaluation verfolgen können. Während den Durchführenden einer Veranstaltung etwa dessen Optimierung vor Augen schwebt, kann für die Financiers die Entscheidungsfindung über den Fortbestand der Veranstaltung bedeutsam sein. Zum zweiten Schritt gehört zudem die Entwicklung, Formulierung und Festlegung von Fragestellungen, welche die Evaluation in Bezug auf den Gegenstand beantworten soll. Darüber hinaus sollen in diesem Ablaufmodell bereits zu Beginn der Evaluation Kriterien bestimmt und begründet werden, mit denen der Gegenstand beschrieben und bewertet wird. In der Regel müssen Kriterien, die sich nicht direkt messen lassen, in beobachtbare Indikatoren übersetzt werden (vgl. Rädiker, 2009). Es empfiehlt sich, bei allen Kriterien Ausprägungen festzulegen, ab denen von „Erfolg“ gesprochen werden kann, das heißt, es sollten also Standards bzw. „Kriterienpunkte“ oder „Kriterienzonen“ (Beywl & Niestroj, 2009, S. 66) für die in Schritt 5 vorzunehmende Bewertung festgelegt werden.

### *3. Schritt: Erhebung planen und Erhebungsinstrumente entwickeln*

Im dritten Schritt wird auf Basis der Ergebnisse des zweiten Schrittes (Zweck, Fragestellungen, Bewertungskriterien) zunächst ein Evaluationsdesign entwickelt und schriftlich fixiert. Darüber hinaus muss über den Einsatz geeigneter und angemessener Evaluationsinstrumente entschieden werden. Falls noch nicht vorhanden, müssen neue Instrumente erarbeitet werden.

#### *4. Schritt: Daten erheben und auswerten*

Der vierte Schritt beginnt mit der Datenerhebung, z.B. mit der Verteilung und Sammlung von Fragebögen, der Durchführung und Protokollierung von Beobachtungen oder der Zusammenstellung und Aufbereitung von Dokumenten. Es schließt sich die Auswertung an, die in der Regel durch Computerprogramme unterstützt wird, etwa bei der Zusammenstellung von Ergebnissen oder deren späterer grafischer Aufbereitung.

#### *5. Schritt: Ergebnisse bewerten und Schlussfolgerungen ziehen*

Die aufbereiteten Ergebnisse werden im fünften Schritt anhand der im ersten Schritt festgelegten Kriterien und Standards bewertet. Zur Bewertung kommen auch Vergleichsdaten in Frage, wie etwa die Vorjahresergebnisse bei wiederholten Programmen. Auf der Grundlage der Bewertungen ziehen die Beteiligten an der Evaluation begründete Schlussfolgerungen. Je nachdem, welche Evaluationszwecke verfolgt wurden, sind größere Entscheidungen, etwa zum Fortbestand eines Programms, oder mehrere kleinere Entscheidungen, etwa hinsichtlich der Verbesserung eines Programms, zu treffen.

#### *6. Schritt: Ergebnisse dokumentieren, kommunizieren und umsetzen*

Im sechsten Schritt werden das Vorgehen, die Evaluationsergebnisse und die Schlussfolgerungen verschriftlicht, um die Transparenz und Verbindlichkeit zu sichern. Es empfiehlt sich, die Ergebnisse und Schlussfolgerungen zu diskutieren und zu kommunizieren, etwa in Sitzungen, Workshops oder per E-Mail. Zum Abschluss des Evaluationsprozesses steht die konkrete Umsetzung der Konsequenzen an, welche bei Selbstevaluationen den Evaluierenden selbst und bei externen Evaluationen in der Regel den Programmverantwortlichen obliegt.<sup>42</sup>

## **2.7 Herausforderungen bei der Evaluation von Weiterbildungsprozessen**

Sowohl im Vorfeld einer Evaluation als auch während und nach einer Evaluation können vielfältige Herausforderungen und Probleme auftreten. Als generelle Herausforderung wird häufig das Problem zwischen einerseits theoretisch formulierten Anforderungen an die Evaluation und andererseits praktischer Realisierungsmög-

---

<sup>42</sup> Habeck und Seitter (2012) beschreiben ein Beispiel, bei dem sie das evaluative Anregungspotenzial einer durchgeführten Studie für die Programmverantwortlichen diskutieren.

lichkeit genannt, denn „wissenschaftlich-methodische Ansprüche an das Evaluationsvorgehen stehen oft im Gegensatz zu einem in der Praxis realisierbaren Vorgehen“ (Häring, 2003, S. 97). In den folgenden Unterkapiteln werden nun konkrete Herausforderungen dargestellt, die bei der Evaluation von Weiterbildungsprozessen auftreten können und die sich zumeist unabhängig vom Grundproblem der Theorie-Praxis-Differenz stellen. Die Darstellung dieser Herausforderungen bereitet die empirische Untersuchung vor, die sich speziell den Herausforderungen der Evaluationspraxis in Weiterbildungsorganisationen widmet. Beispielsweise ist es möglich, die einzelnen Herausforderungen in den geplanten Interviews mit Evaluationsverantwortlichen abzufragen.

### 2.7.1 Widerstände gegen Evaluation

Mit am häufigsten diskutiert wird das Thema Widerstand gegen Evaluation, das im englischen Sprachraum unter dem Begriff „resistance“ aufgegriffen wird und das mit dem Thema Akzeptanz von Evaluation gekoppelt ist. Der Grund für die wichtige Bedeutung der Themen Widerstand und Akzeptanz ist sicherlich darin zu sehen, dass bei Evaluationen immer auch ein Kontrollaspekt mitschwingt, der von den Beteiligten teilweise auf die eigene *Person* bezogen wird, wenngleich es meist nur um die eigene *Arbeit* geht. Beywl (2001, S. 21) beschreibt, welche Hindernisse bei der Einführung von Evaluation in der Jugendverbandsarbeit auftreten können. Er benennt unter anderem persönliche Ängste und Befürchtungen, die bei den Evaluierten durch die Evaluation – auch bei besonders erfahrenen Personen – auftreten können. Beywl macht darauf aufmerksam, dass insbesondere solche Evaluationen als Bedrohung erlebt werden können, die in Zeiten knapper Finanzmittel dazu instrumentalisiert werden, unliebsame Finanzierungs- oder Personalentscheidungen zu begründen (vgl. ebd., S. 23).

Was bedeutet Widerstand im Kontext von Evaluationen im Detail? Nach Zuschlag (1987, S. 77) kann man unter Widerstand gegen Evaluationsmaßnahmen alle (Re-)Aktionen, Einstellungen oder Empfindungen einer Person verstehen, die ihre selbstgesetzten oder von außen an sie herangetragene Anforderungen bremsen oder blockieren. Widerstand kann sich dabei aktiv in Form einer Handlung (z.B. im Kontext der Weiterbildung als Beschwerde eines Dozenten über die Evaluation) oder passiv in Form der Nicht-Handlung (z.B. Nicht-Ausfüllen des Teilnehmendenfragebogens) zeigen. Taut und Brauns (2003, S. 48) sehen „resistance“ als ein menschliches Verhalten, um den Status quo aufrecht zu erhalten. Ihrem Verständnis

nach kann sich Widerstand gegen die Evaluation selbst richten (z.B. gegen den verwendeten Ansatz oder die Interpretation der Ergebnisse), gegen den Kontext, in dem die Evaluation stattfindet (z.B. gegen die Eigenschaften des Programms, das evaluiert wird), oder gegen die Personen und Personengruppen, die an der Evaluation beteiligt sind (z.B. gegen deren Erwartungen und Werturteile).

Zuschlag (1987) entfaltet im Kontext der beruflichen Aus- und Weiterbildung ein differenziertes Modell zu Widerständen gegen Evaluationsmaßnahmen, das drei Dimensionen unterscheidet, die in Wechselbeziehung zueinander stehen. Die erste Dimension bezieht sich auf den Lernprozess, während die zweite Dimension auf den Evaluationsprozess gerichtet ist. In Abb. 21 sind die einzelnen Komponenten des Lern- und Evaluationsprozesses aufgeführt, aus denen Widerstand herrühren, die Widerstand verursachen oder auf die sich Widerstand beziehen kann. Die Komponenten lassen sich unterteilen in einerseits unmittelbare, also direkt auf den Lern- und Evaluationsprozess bezogene Komponenten, und andererseits mittelbare Komponenten, die nicht direkt auf den Lern- und Evaluationsprozess bezogen sind.

**Abb. 21: Quellen von Widerständen in Lernprozessen und Evaluationsprozessen (nach Zuschlag, 1987, S. 78)**

	Am Lernprozess beteiligte Komponenten	Am Evaluationsprozess beteiligte Komponenten
unmittelbar	<ul style="list-style-type: none"> <li>• Lernende, Lerngruppe</li> <li>• Lehrende</li> <li>• Lernstoff</li> <li>• Lehrmethode</li> <li>• Lern-/Lehrsituation</li> </ul>	<ul style="list-style-type: none"> <li>• Evaluierter Person(en)</li> <li>• Evaluator</li> <li>• Evaluations-Inhalt</li> <li>• Evaluations-Methode</li> <li>• Evaluations-Situation</li> </ul>
mittelbar	<ul style="list-style-type: none"> <li>• Auftraggebende/Arbeitgebende</li> <li>• Lern-/Lehrauftrag</li> <li>• Betriebsrat</li> <li>• Konsequenzen des Lernens (Lernergebnis)</li> </ul>	<ul style="list-style-type: none"> <li>• Auftraggebende/Arbeitgebende</li> <li>• Evaluations-Auftrag</li> <li>• Betriebsrat</li> <li>• Konsequenzen der Evaluation (Evaluationsergebnis)</li> </ul>

Die dritte Dimension des Modells von Zuschlag berücksichtigt das Ausmaß des Widerstands und meint konkret, wie viele der Komponenten vom Widerstand betroffen sind; also ob es sich um generelle Widerstände, um Widerstände bezogen auf mehrere Komponenten oder nur auf eine Komponente bezogen handelt.

Es stellt sich die Frage, welche Erklärungen für Widerstände gegen Evaluationen herangezogen werden können. Taut und Brauns (2003, S. 249-251) haben hierzu verschiedene in der Literatur vorzufindende psychologische Kernkonzepte zusammengestellt: *Attitudes toward evaluation* – Negative Einstellungen können insbesondere im Kontext der Verwendung von Evaluationsergebnissen Widerstand erzeugen.



*Maintaining the status quo* – Hierunter ist die Verteidigung des aktuellen Stands zu verstehen, z.B. aus Angst vor Kritik und der Angst Macht einzubüßen. *Power* – Evaluation kann auch Machtunterschiede ansprechen und unterschiedliche Interessen offenlegen und benötigt daher immer eine Form der Konfliktlösung. *Reactance* – Allein die Aussicht, durch die Evaluation würde die Wahlfreiheit eingeschränkt, löst Widerstände aus.

Eng verbunden mit dem Problem des Widerstands ist die Belastung der Teilnehmenden durch die Evaluation, die man sich leicht vor Augen führen kann, wenn beispielsweise Teilnehmende einer Tagung sechsmal den gleichen Evaluationsbogen für sechs unterschiedliche Workshops ausfüllen sollen. Die Belastung der Teilnehmenden wird zwar sicherlich nicht so prominent in der Literatur diskutiert wie das Thema Widerstand, dennoch findet das Thema im Evaluationsstandard „D1 Angemessene Verfahren“ der DeGEval Berücksichtigung. Dort heißt es: „Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen“ (DeGEval, 2008, S. 11). Reischmann (2006, S. 101) weist darüber hinaus auch auf die Belastung der evaluierten Lehrenden hin, wenn er fragt: „Kann man Ehrenamtlichen, die aus persönlichem Engagement und bestenfalls gegen ein Taschengeld Kurse halten, zumuten, sich einer Kontrolle ihrer Vermittlungsqualität auszusetzen?“

### 2.7.2 Wirkungsmessung und -nachweis

Insbesondere im Rahmen der beruflichen und betrieblichen Bildung aber auch in der Allgemeinen Erwachsenenbildung werden häufig Herausforderungen bezüglich der Erfassung von Wirkungen einer Bildungsmaßnahme thematisiert und es haben sich zahlreiche Autoren/innen mit diesem Problem beschäftigt (etwa Beywl & Balzer, 2008, S. 10; Holla, 2002, S. 127-129; Reischmann, 2006, S. 95-99). Die unterschiedlichen Facetten der Problematik spiegeln sich vor allem in den folgenden Aspekten wider:

- Die Wirkung einer Bildungsmaßnahme kann sehr vielseitig sein. Neben den intendierten Wirkungen sind zahlreiche weitere Effekte denkbar und es lassen sich adhoc, kurzfristige, mittelfristige und langfristige Wirkungen unterscheiden. (Weiter-)Bildung ist ein reflexives Gut, das in großen Teilen durch die Lernenden selbst hergestellt wird (vgl. Abschnitt 3.1). Dementsprechend können die

Wirkungen auch individuell sehr verschieden ausfallen und sind stark von verschiedenen Einflussfaktoren wie etwa Motivation, Themenrelevanz und Vorerfahrung abhängig. Zudem lassen sich manifeste Wirkungen leichter intersubjektiv überprüfen als latente, subjektive Wirkungen.

- Es ist ein durchweg schwieriges Unterfangen, Wirkungen auf eine konkrete Bildungsmaßnahme zurückzuführen, weil in der Regel Zeit vergeht, bis sich Wirkungen einstellen können und die Teilnehmenden neben der Bildungsmaßnahme zahlreichen anderen Einflussfaktoren ausgesetzt sind. Es stellt sich also ein kaum überwindbares Kausalitätsproblem. (Vgl. Zech, 2008, S. 91)
- Um Wirkungen nachzuweisen, benötigt man ein Konzept darüber, *was* eine Bildungsmaßnahme bewirken und *wie* – also mit welchen Mechanismen – sie überhaupt wirken soll. Wirkungsmessungen benötigen zudem aufwändige Instrumente und Designs. Für den wissenschaftlichen Nachweis von Wirkungen sind Experimentaldesigns mit Kontrollgruppen angeraten, welche jedoch sehr aufwändig sind und sich nicht nach Belieben in den Alltag von Weiterbildungsorganisationen integrieren lassen. (Vgl. Beywl & Balzer, 2008, S. 10; Häring, 2003, S. 252)
- Häufig werden Einschätzungen von und Meinungen über Wirkungen als Wirkungsnachweise ausgegeben (vgl. Beywl & Balzer, 2008, S. 10).

### 2.7.3 Unklare Gegenstandsziele

Wenn man unter Evaluation die Bewertung von Daten auf der Grundlage von Zielen oder Soll-Größen versteht, wie es beispielsweise Zech (2008, S. 90) betont, dann besteht eine weitere Herausforderung darin, dass für die Evaluation von Weiterbildungsprozessen auch Bewertungsmaßstäbe und Standards vorliegen müssen. Sofern unklar ist, welche Ziele eine einzelne Weiterbildungsmaßnahme, ein Veranstaltungs-Curriculum oder das gesamte Weiterbildungsangebot verfolgt, lässt sich deren Gelingen nur schwerlich an der Zielerreichung ablesen.

### 2.7.4 Auswahl und Definition von Indikatoren und Kriterien

Im Rahmen von Evaluationen werden in der Regel Kriterien und Indikatoren definiert, anhand derer der Gegenstand bewertet wird. Dabei stellt sich zunächst das Problem, dass eine Evaluation niemals den Evaluationsgegenstand in seiner Gänze erfassen kann und daher eine Auswahl getroffen werden muss, mit welchen Kriteri-

en und Indikatoren der Gegenstand beleuchtet werden soll. Daraus ergibt sich die Herausforderung, dass sich jede Wirkung, jeder Erfolg einer Weiterbildungsmaßnahme ad absurdum führen lässt, wenn die Indikatoren zur Bewertung von Wirkung und Erfolg ausgetauscht werden. Was eben noch als positiv bewertet wurde, wird durch den Wechsel der Indikatoren jetzt als negativ beurteilt; während der Verkaufsumsatz nach der Schulung stark angestiegen ist, kann zugleich die persönliche Zufriedenheit mit der Art des Verkaufes stark gesunken sein. Dieses Problem bezeichnet Reischmann (2006, S. 101) als „Indikatorenschaukel“. Weil die Festlegung der Kriterien und Indikatoren für die Evaluation von so entscheidender Bedeutung ist, wird häufig auch diskutiert, wer das Recht hat, über diese zu entscheiden bzw. wer an der Festlegung alles beteiligt ist.

### **2.7.5 Standardisierung und Vereinheitlichung von Instrumenten**

Bezüglich der verwendeten Instrumente finden sich in der Literatur zur Evaluation häufig nur allgemein formulierte Herausforderungen, wie etwa dass es schwierig ist und Kompetenzen erfordert, diese zu entwerfen, und dass es wichtig ist, diese an die aktuelle Situation anzupassen. Wesseler (2010) macht hingegen auf ein ganz konkretes Problem in der allgemeinen Weiterbildung aufmerksam:

„Eine besondere Schwierigkeit der Programmevaluation im Bereich der Erwachsenenbildung liegt in der oft sehr großen Heterogenität der Teilnehmergruppen. Die Verwendung eines einzigen Instruments führt zu ganz unterschiedlichen Reaktionen und Deutungen von Seiten der Teilnehmer, so dass nicht nur die Vergleichbarkeit zwischen einzelnen Gruppen bzw. Kursen problematisch wird, sondern auch insgesamt die Zuverlässigkeit der erhobenen Daten in Zweifel gezogen werden muss, insbesondere dann, wenn Mittelwerte gebildet werden.“ (S. 1043)

Wesseler spricht hier das Spannungsfeld der Standardisierung und Vereinheitlichung von Instrumenten einerseits und der Spezialisierung von Instrumenten andererseits an. Dieses Problem stellt sich sowohl für sehr heterogene Kurse (etwa Sportkurse vs. EDV-Kurse) als auch für Kurse mit sehr heterogenen Teilnehmendengruppen. Hinzu kommt das Problem der Aggregierung von Daten, denn es stellt sich immer die Frage, inwieweit es (für wen) hilfreich und nutzbringend ist, Mittelwerte für Kurse, Kursreihen oder gar mehrere Jahre zu berechnen und dabei die konkreten Häufigkeitsverteilungen außer Acht zu lassen.

### **2.7.6 Aufwands- und Ressourcenprobleme**

Auf einer anderen Ebene liegt die Problematik der Ressourcen. Diese können zum einen nicht ausreichend für die Durchführung von Evaluationen vorhanden sein und zum anderen können Evaluationen als zu aufwändig angesehen werden, was beides gleichermaßen Widerstand hervorrufen kann. Häring (2003, S. 247-250) stellt die Ergebnisse einer Interviewstudie unter 22 Personalentwicklern/innen von Unternehmen vor, in der die Autorin ressourcenbedingte Hinderungsgründe für die Durchführung von Evaluationen herausarbeitet und die drei Faktoren Zeitaufwand, Kosten und Personalressourcen benennt. Geringe Ressourcen führten laut der Untersuchung zu einer Einschränkung der Methodenwahl und des Umfangs, indem nur einfache Instrumente und Verfahren ausgewählt und nicht alle Trainingsmaßnahmen evaluiert werden.

### **2.7.7 Defizite bezüglich Nutzen und Nutzung**

Es gibt eine umfangreiche Diskussion über den Nutzen von Evaluationen und die Nutzung von Evaluationsergebnissen (z.B. Stockbauer, 2000). Dahinter verbirgt sich der Anspruch, dass Evaluationen nicht nur einem Selbstzweck dienen, sondern einen Nutzen erbringen sollen. Die Herausforderung besteht also darin, diesen Anspruch auch einzulösen und dafür zu sorgen, dass Evaluation einen Nutzen hat *und* die Evaluationsergebnisse genutzt werden. Reischmann (2006, S. 102) stellt in seinem Buch „Weitebildung-Evaluation“ in einer Überschrift die provokative Frage „Nutzt Evaluation überhaupt etwas?“ und berichtet anschließend von Zweifeln, die an der Sinnhaftigkeit und dem Nutzen von Evaluation im Weiterbildungsbereich auftreten können, wenn man auf Beispiele folgenloser Evaluation schaut oder die teilweise immensen Kosten gegenüberstellt. Die Diskussion um den Nutzen steht also in einem engen Zusammenhang zu Aufwand und Ressourcen, was in der einfachen Frage mündet, in welchem Verhältnis Aufwand und Ertrag einer Evaluation stehen.

### **2.7.8 Fehlende Evaluationskompetenzen**

„Evaluieren kann jeder“ ist eine weit verbreitete Meinung, die Beywl und Balzer (2008) zu einem von zehn großen Irrtümern bezüglich Evaluationen in der Weiterbildung zählen: „Doch selbst, dass man zum Beispiel Statistik, Psychologie und Soziologie studiert hat, heißt nicht automatisch, dass man die Wissenschaft und Kunst der Evaluation beherrscht. Dies erfordert mehrjährige Berufspraxis und weiterbil-

dende Studien, analog zu Supervision, Coaching oder Organisationsentwicklung“ (S. 11).

Fehlende Evaluationskompetenzen können für die Praxis eine große Herausforderung darstellen. So bemerkt Reischmann (2006, S. 102), dass bereits die Entwicklung und Anwendung von Instrumenten methodisches Fachwissen erfordert und die Designs aufwändig sein können. Nach Reischmann verfügt jedoch das Weiterbildungspersonal nur selten über die Fähigkeit, dies fachgerecht zu leisten. Will et al. (1987a, S. 11) weisen in diesem Zusammenhang darauf hin, dass Evaluationen im Weiterbildungsbereich zudem das Gefühl der Überforderung hervorrufen können.



### **3 Die Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW)**

Nachdem im vorhergehenden Kapitel die Grundlagen der Evaluation von Weiterbildungsprozessen erörtert wurden, legt das dritte Kapitel einen weiteren Grundstein für den empirischen Teil der vorliegenden Arbeit: In den Abschnitten 3.1 und 3.2 werden zunächst die Grundzüge der Lernerorientierten Qualitätstestierung vorgestellt sowie die Frage beantwortet, welchen Stellenwert hierbei die Evaluation von Bildungsprozessen einnimmt. Als Vorbereitung für die Erarbeitung der Fragestellungen schließt sich in Abschnitt 3.3 die Vorstellung existierender Forschungsergebnisse zur Evaluation von Weiterbildungsprozessen an. Hierzu werden zum einen Studienergebnisse berichtet, die sich direkt auf die Lernerorientierte Qualitätstestierung beziehen, zum anderen aber auch solche Ergebnisse berücksichtigt, welche in anderen Kontexten entstanden sind.

#### **3.1 Die Grundzüge des LQW-Modells**

Die „Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW)“ ist ein Qualitätsmanagementverfahren, das von 2000 bis 2005 im Rahmen mehrerer Forschungs- und Entwicklungsprojekte speziell für die Weiterbildung entwickelt wurde. LQW ist inzwischen das in Deutschland am weitesten verbreitete Qualitätsverfahren für Weiterbildungsorganisationen<sup>43</sup> und zu Beginn des Jahres 2012 setzen gut 450 Organisationen aus unterschiedlichsten Bereichen der Weiterbildung das Qualitätsmanagementverfahren bundeweit ein. Hinzu kommen knapp 100 Organisationen in Österreich und seit wenigen Jahren auch mehrere Organisationen in Bulgarien, Litauen und Polen (vgl. [www.artset-lqw.de](http://www.artset-lqw.de)).

Dem LQW-Modell liegt die zentrale Überlegung zugrunde, dass Bildung ein reflexives Gut ist, dessen Endprodukt „Lernerfolg“ nicht von der Weiterbildungsorganisation hergestellt werden kann, sondern von den Lernenden selbstständig erarbeitet werden muss (vgl. Zech & Braucks, 2004, S. 11-12). Aus diesem Grund bieten Weiterbildungsorganisationen auch keine Bildung an, sondern die „infrastrukturell

---

43 In dieser Arbeit wird dem Begriff „Organisation“ gegenüber den Bezeichnungen „Einrichtung“ und „Institution“ zumeist Vorrang gegeben, denn wie Körber (2002) herausarbeitet, beziehen sich letztere (verwaltungstechnische) Begriffe häufig nur auf staatliche Organisationen, deren Prototyp die Volkshochschule darstellt. Das Spektrum der Organisationen, die LQW einsetzen, umfasst jedoch weit mehr als nur staatliche Einrichtungen.

und interpersonell unterstützte Möglichkeit, sich zu bilden“ (ebd., S. 14). Weiterbildungsorganisationen stellen also nur die Möglichkeit für Bildung, genauer gesagt: die Bedingungen für die Möglichkeit von Bildung, bereit. Die Lernerorientierte Qualitätstestierung greift diesen Grundgedanken auf, indem sie sich nicht auf die *Qualität von Bildung*, sondern auf die *Qualität der Organisation der Bedingungen von Bildung* konzentriert (vgl. Zech et al., 2006, S. 24-28). Zur Verbesserung dieser Qualität setzt LQW primär auf die Steigerung der Lernfähigkeit der Organisation, und zwar vor allem durch die Steigerung ihrer Fähigkeiten zur Reflexivität, was beispielsweise in dem Titel der Einführungspublikation zu LQW „Qualität durch Reflexivität“ (Zech, 2004) zum Ausdruck kommt. Im LQW-Modell wird der Reflexivität explizit der Vorzug gegenüber einer Formalisierung gegeben, da die bloße Orientierung an Schemata und Checklisten die Qualitätsentwicklung und das Lernen von Organisationen auch behindern könne. Zech (2006) merkt hierzu an:

„LQW setzt deshalb auf Reflexivität vor Formalität. Wichtig ist, dass die Organisation herausarbeitet, warum ein bestimmter Prozess in welcher Weise die Bedingungen für gelungenes Lernen der Teilnehmenden verbessert. Hierfür ist eine Verständigung über pädagogische Leitziele notwendig. Wenn dies klar ist, dann kann man auch entscheiden, ob eine Formalisierung diesem Prozess förderlich ist oder ob sie die Qualität sogar einschränkt.“ (S. 9)

Um diesen Grundgedanken und den damit verbundenen Ansprüchen gerecht zu werden, folgt die Lernerorientierte Qualitätstestierung einem speziellen Aufbau und sieht einen bestimmten Ablauf vor. In Abb. 22 ist der Lernerorientierte Qualitätsentwicklungsprozess schematisch dargestellt. Wie der Grafik zu entnehmen ist, umfasst das LQW-Modell insgesamt elf Qualitätsbereiche, die sich bei Bedarf um weitere organisationsspezifische, optionale Bereiche erweitern lassen.



Abb. 22: Der Lernerorientierte Qualitätstestierungsprozess (Quelle: Zech, 2006, S. 13)



Der erste Qualitätsbereich zum „Leitbild“ steht ganz links in der Grafik und nimmt eine Sonderstellung ein, denn eine Weiterbildungsorganisation beginnt ihren Qualitätsprozess damit, ein Leitbild zu entwickeln, in dem die Identität der Organisation zum Ausdruck kommt. Dieses Leitbild muss auch zwingend eine sogenannte „Definition gelungenen Lernens“ enthalten. Diese Definition gelungenen Lernens (DgL) stellt das Alleinstellungsmerkmal von LQW gegenüber anderen Qualitätsverfahren in der Weiterbildungsbranche dar. Mithilfe der DgL soll die Paradoxie, dass Qualitätsentwicklung einerseits die Lernprozesse und den Lernerfolg verbessern soll, andererseits die Bildungsorganisation aber keinen mittelbaren Zugriff auf das Lernen ihrer Teilnehmenden hat, überwunden werden (vgl. Zech & Braucks, 2004, S. 17). In der Definition gelungenen Lernens hält die Weiterbildungsorganisation ihr pädagogisches Selbstverständnis fest und beschreibt, wie ein optimal gelungener Lernprozess aussieht und was diesen kennzeichnet. Dabei versetzt sich die Organisation in die Perspektive ihrer Teilnehmenden und betrachtet sich selbst und die Lernprozesse aus dieser Position heraus. Bereits die Erarbeitung der Definition erfordert und trainiert also die Reflexivität der Organisation. Wichtig zu betonen ist, dass sich die Definition gelungenen Lernens nicht an realen Teilnehmenden orientiert, sondern als „idealtypische Konstruktion des besten denkbaren Falls“ (Zech & Braucks, 2004, S. 17) zu verstehen ist, welche die Qualitätsentwicklung zusammen mit dem Leitbild anleitet und ihren roten Faden darstellt. Hervorzuheben ist dabei auch die Bedeutung des bewusst gewählten Wortes „gelingen“, die bereits deutlich wird, wenn man rein sprachlich einen *gelungenen* Lernprozess mit einem *erfolgreichen* Lernprozess

vergleicht. Während bei einem gelungenen Lernprozess die Bedeutungen von gelingen, glücken und gedeihen sowie von Innerlichem, Zufriedenheit und Erfüllung mit-schwingen, verbindet man mit einem erfolgreichen Lernprozess stärker etwas Äu-ßerliches, einen Status und die reine Zielerreichung (vgl. Zech et al., 2006, S. 34-35).

In allen elf Qualitätsbereichen von LQW sind Mindestanforderungen definiert, deren Erfüllung eine Weiterbildungsorganisation für eine erfolgreiche Qualitätstestierung nachweisen muss. Diese Anforderungen lassen sich nicht checklistenartig abhaken, sondern sie sind so formuliert, dass sie zwar hinsichtlich ihrer Erfüllung abgeprüft werden können und eine organisationsübergreifende Vergleichbarkeit gewährleisten, zugleich aber die organisationsinterne Reflexivität fördern. Die Anforderungen zwingen die Organisationen dazu, über ihr Handeln in den einzelnen Qualitätsbereichen zu reflektieren und dieses zu begründen. Beispielsweise heißt es im Qualitätsbereich „Führung“: „Instrumente und Verfahren der internen Kommunikation und Information sind beschrieben und eingeführt“ (Zech, 2006, S. 72) und nicht „Die Einrichtung hält einmal pro Woche eine Teamsitzung ab“. Fast alle Qualitätsbereiche enthalten die Mindestanforderung, dass die jeweiligen Abläufe und Tätigkeiten in Bezug auf das Leitbild und die Definition gelungenen Lernens begründet werden. Hierdurch wird gewährleistet, dass das organisationsspezifische Leitbild den roten Faden der Qualitätsentwicklung bildet und alle Maßnahmen dazu beitragen, sich der Idealvorstellung gelungenen Lernens in der täglichen Praxis weiter anzunähern.

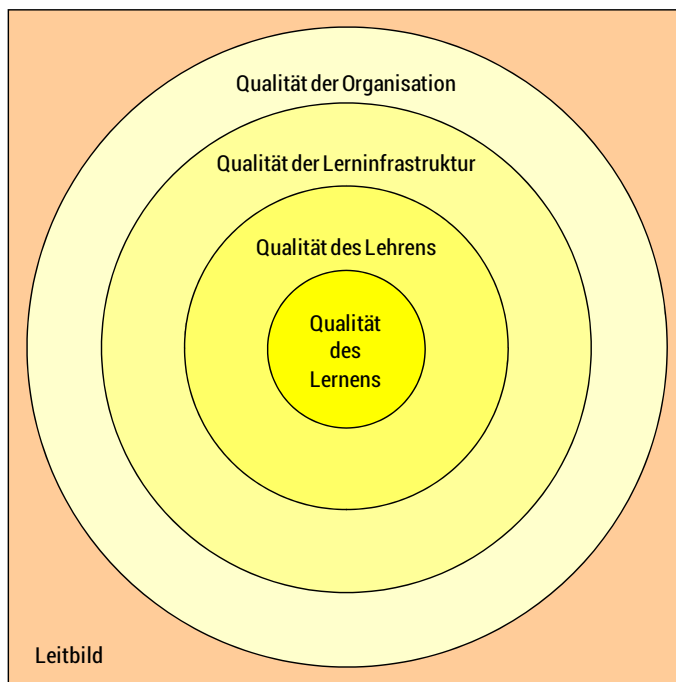
Maximal dreizehn Monate nach ihrer Anmeldung zur Ersttestierung muss eine Organisation einen Selbstreport vorlegen, der ca. 80 Seiten umfasst. In diesem Selbstreport stellt sich die Organisation kurz dar und beschreibt selbsterklärend und mit Nachweisen untermauert, wie sie die einzelnen Mindestanforderungen erfüllt. Zwei Gutachter/innen prüfen den Selbstreport hinsichtlich der Erfüllung der Mindestanforderungen, welche für jeden Qualitätsbereich im LQW-Modell definiert sind. Sie fertigen ein Gutachten an, das in knapper Form über das Ergebnis dieser Prüfung informiert. Den weit größeren Anteil des Gutachtens nehmen jedoch Kommentare und Anmerkungen ein, die zahlreiche „Hinweise, Vorschläge und Ideen für die weitere Organisations- und Qualitätsentwicklung“ (ebd., S. 51) enthalten und in denen die Gutachter/innen ihre Außenwahrnehmung (ohne Wertungen, sondern im Sinne einer Wiederbeschreibung) an die Organisation zurückspiegeln. Durch dieses Verfahren erhält die Weiterbildungseinrichtung auch Hinweise auf Entwicklungspotenziale und Chancen, was LQW einen Doppelcharakter von *Qualitätstestierung* und *Qualitätsentwicklung* verleiht. Nach der Übersendung des Gutachtens an die Organisation

nimmt der Erstgutachter bzw. die Erstgutachterin eine Visitation der Organisation vor, auf der Fragen zum Gutachten besprochen und die Erfüllung noch offen gebliebener Mindestanforderungen nachgewiesen werden muss. In einem Abschlussworkshop, der in Abb. 22 ganz rechts am Ende des Testierungsprozesses steht, verständigen sich Organisation und Erstgutachter/in über strategische Entwicklungsziele, welche die Organisation im elften Qualitätsbereich erarbeitet hat und zu denen sich die Organisation für die Fortsetzung ihrer Qualitätsarbeit verpflichtet.

Nach erfolgreicher Testierung erhält die Organisation ein Testat zusammen mit einer Keramikfliese aus einem Netzwerkbild aller testierten Organisationen, das die Gemeinschaft der Weiterbildungsorganisationen und ihre gemeinsame Qualitätsarbeit symbolisiert. Das Testat hat eine Gültigkeit von vier Jahren. Nach vier Jahren muss sich die Organisation zur Retestierung angemeldet haben und ihren überarbeiteten Selbstreport zur Begutachtung abgeben. Der aktualisierte Selbstreport spiegelt den aktuellen Stand zu den elf Qualitätsbereichen wider und beschreibt die Erfüllung der vereinbarten strategischen Entwicklungsziele. Der Qualitätsprozess endet also nicht mit der Testierung, denn wie Meyer (2006) es in dem Titel eines Erfahrungsberichtes formuliert: „Nach der Testierung ist vor der (Re)Testierung“. Vielmehr soll LQW einen kontinuierlichen Organisationsentwicklungsprozess befördern, der sich sowohl in stetig wachsender als auch veränderter Qualität niederschlägt (vgl. Zech, 2006, S. 43).

Zwar mag die schematische Darstellung des LQW-Modells einen linearen Prozess suggerieren, doch ist das Modell von seiner Idee her in konzentrischen Ringen aufgebaut (vgl. Abb. 23). Mittelpunkt aller Bemühungen stellt die Qualität des Lernens dar, welche von der Qualität des Lehrens, der Qualität der Lerninfrastruktur und der Qualität der Organisation umringt ist. Je weiter außen ein Ring steht, desto weiter ist er vom Lerngeschehen entfernt. Umrahmt werden diese Qualitätsringe vom Leitbild der Organisation.

Abb. 23: Die Kontexte des Lernens (Quelle: ArtSet, 2006, S. 5)



Die folgenden sechs Punkte, entnommen aus dem „Leitfaden für die Praxis“ (Zech, 2006), fassen noch einmal die Grundgedanken der Lernerorientierten Qualitätstestierung kondensiert zusammen:

- „1. Die Lernenden als ‚kundige‘ Produzenten von Bildung stehen im Mittelpunkt aller Qualitätsbemühungen. Auf sie ist die Qualitätsentwicklung der Organisation und mithin das Testierungsverfahren ausgerichtet.
2. Es geht nicht nur um Qualitätssicherung, sondern die Qualität der Organisationen wird in einem ständigen Prozess *weiterentwickelt* – ausgehend von den sich verändernden Umweltanforderungen. Das Lernen der Organisation ist dabei die Basis der Verbesserung des Lernens der Teilnehmenden.
3. Bildung ist ein reflexives ‚Erfahrungsgut‘; die Verbesserung der organisationalen Bildungsbedingungen muss daher auch ein reflexiver Prozess sein. Qualitätsverbesserung in der Bildungsbranche kann nicht durch technokratische Formalisierungen gefördert werden.
4. Es handelt sich bei LQW daher nicht nur um ein externes Begutachtungsverfahren, vielmehr werden die Entwicklungspotenziale der Organisationen berücksichtigt und gefördert. LQW unterstützt die Lernprozesse der Weiterbildungseinrichtungen und prüft nicht fremdgesetzte Standards ab.

5. Das Modell ist organisationstypübergreifend sowie für große und kleine Organisationen gleichermaßen anwendbar, d.h. jede Organisation kann LQW an ihre besonderen Bedingungen anpassen. Das Modell ist also selbst lernfähig und anwendungsflexibel einsetzbar.
6. Im LQW-Netzwerk dient das Qualitätsmodell der Vergleichbarkeit der Weiterbildungsorganisationen untereinander; hierdurch wird Organisationslernen über wechselseitige Beratung gefördert.“ (S. 12)

Wie diese sechs Charakteristika deutlich zeigen, versteht sich die Lernerorientierte Qualitätstestierung nicht als eine bloße *Zertifizierung*, in der von einer dafür legitimierten Stelle durch eine externe Prüfung die Qualität einer Weiterbildungsorganisation bestätigt wird. LQW stellt vielmehr ein *Testierungsverfahren* dar, welches neben der reinen Testierung vor allem auf Förderung von Reflexivität setzt und dabei auf planvolle Weise Elemente der Selbstevaluation mit externer Qualitätsevaluation kombiniert. Wie Zech und Tödt (2012, S. VIII) hervorheben, wird im Unterschied zur Zertifizierung bei der Testierung auch die Gegenstandsangemessenheit und Plausibilität von Prozessen und Verfahren begutachtet und es wird die Fähigkeit zu Reflexion und kontinuierlichem Organisationslernen bestätigt. Der Begriff der Testierung im LQW-Modell soll zudem die Selbstbestimmung der inhaltlichen Standards durch die professionell Handelnden betonen (vgl. ebd.).

### **3.2 Die Evaluation von Bildungsprozessen im Rahmen des LQW-Modells**

Der fünfte Qualitätsbereich von LQW befasst sich mit der Evaluation von Bildungsprozessen<sup>44</sup>, denn „ohne Rückmeldungen kann keine Bildungsorganisation empirische Aussagen über den Erfolg der geleisteten Arbeit treffen“ (Zech et al., 2006, S. 40). In Abb. 24 ist der fünfte Qualitätsbereich zur Evaluation der Bildungsprozesse dargestellt, wie er auch im LQW-Leitfaden abgedruckt ist.

---

44 Der LQW-Testierungsprozess als Ganzes kann in Ansätzen auch als Evaluationsprozess aufgefasst werden. Dieser Evaluationsprozess bezieht sich jedoch nicht (nur) auf die Mikroebene der Bildungsprozesse, sondern primär auf die Mesoebene der Organisation (vgl. Abschnitt 2.2).

**Abb. 24: Der fünfte Qualitätsbereich von LQW zur „Evaluation der Bildungsprozesse“**  
(Quelle: Zech, 2006, S. 68)

<b>Evaluation der Bildungsprozesse</b>		
Evaluation von Bildungsprozessen bedeutet, dass die durchgeführte Bildungsarbeit regelmäßig mit geeigneten Instrumenten geprüft und bewertet wird. Maßstabsbildend zur Bewertung sind der Lernerfolg, die Zufriedenheit der Teilnehmenden und ggf. der Auftraggeber sowie die Realisierung des eigenen institutionellen Anspruchs. Auch die Einschätzung der Lehrenden sollte Teil der Evaluation sein.		
Spezifikationen	Anforderungen	Nachweismöglichkeiten
Rückmeldungen von Lernenden über: <ul style="list-style-type: none"> <li>• Lernerfolge,</li> <li>• Zufriedenheit,</li> <li>• Lehr-Lern-Prozesse,</li> <li>• Infrastruktur und</li> <li>• die Arbeit der Lehrenden</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Gegenstände,</b></li> <li>• <b>Verfahren,</b></li> <li>• <b>Rhythmus und Umfang der Evaluation sind beschrieben und in Bezug auf das Leitbild und die Definition gelungenen Lernens begründet.</b></li> </ul>	Evaluationsinstrumente  dokumentierte Evaluationsergebnisse  Protokolle über Evaluationskonferenzen  etc.
Rückmeldungen von Lehrenden	<b>Die Analysen werden bewertet, Konsequenzen aus den Evaluationsergebnissen werden gezogen.</b>	
Rückmeldungen von Auftraggebern	<b>Die Lehrenden werden über die Evaluationsergebnisse informiert.</b>	

Der fünfte Qualitätsbereich folgt dem einheitlichen Aufbau aller Qualitätsbereiche der Lernerorientierten Qualitätstestierung: Im Kopf enthält er zunächst eine Definition dessen, was unter Evaluation der Bildungsprozesse im Rahmen von LQW verstanden wird, um im Bearbeitungs- und Begutachtungsprozess ein gemeinsames und einheitliches Verständnis sicherzustellen.<sup>45</sup> Die linke Spalte enthält Spezifikationen, welche die Definition verdeutlichen und konkretisieren sollen. Die Spezifikationen sind nicht verpflichtend, sondern es handelt sich um Anregungen, welche die Organisation bei der Gestaltung und der Arbeit an ihrer Evaluation der Bildungsprozesse unterstützen sollen. In der mittleren Spalte finden sich fett gedruckt die Mindestanforderungen an die Evaluation der Bildungsprozesse, deren Erfüllung jede Organisation im Selbstreport nachweisen muss. So muss sie beschreiben, welche Gegenstände mit welchen Verfahren in welchem Rhythmus und Umfang evaluiert werden und das gewählte Vorgehen in Bezug auf das Leitbild und die Definition gelungenen Lernens begründen. Darüber hinaus müssen die Analysen bewertet werden, Konsequenzen gezogen und die Lehrenden über die Ergebnisse informiert werden. Die rechte Spalte enthält Beispiele für Nachweise, mit denen die Organisation die Beschreibungen ihrer Evaluationspraxis im Selbstreport belegen kann. Der fertige Selbstreport einer Organisation enthält einen etwa drei- bis fünfseitigen Text,

<sup>45</sup> Die Definition wurde bereits ausführlich in Abschnitt 2.2 vorgestellt.

in dem – häufig gegliedert nach den Mindestanforderungen – die aktuelle Evaluationspraxis beschrieben wird.

### **3.3 Forschungsergebnisse zur Evaluation von Weiterbildungsprozessen**

Es liegen einige Studien über das Qualitätsmanagementverfahren LQW vor, welche Aussagen über die Bedeutung des fünften Qualitätsbereichs zur Evaluation der Bildungsprozesse, insbesondere im Verhältnis zu anderen Qualitätsbereichen, erlauben. Darüber hinaus liefern die Studien Hinweise auf Wirkungen, die Organisationen im Bereich der Evaluation durch die Anwendung von LQW erzielen konnten. Hartz et al. (2007, S. 110) weisen in einer groß angelegten Wirkungsstudie mit zwei Messzeitpunkten unter knapp 200 Organisationen nach, dass Weiterbildungseinrichtungen durch die Implementierung von LQW<sup>46</sup> im Bereich der Evaluation deutlich profitieren: So nehmen jeweils etwa 80% der Organisationen Wirkungen bezüglich der Evaluation der Kurse und der Nutzung der Evaluationsergebnisse wahr, während die Wirkungen in den meisten anderen Bereichen niedriger ausfallen. Die Untersuchung hat zudem gezeigt, dass der Bereich Evaluation auf Platz 2 der wahrgenommenen Entwicklungsbedarfe steht. Dementsprechend haben auch mehr als 85% der Organisationen das Wissen über Evaluation für die Qualitätsentwicklung als wichtig eingeschätzt und knapp 15% der Organisationen gaben im Nachhinein an, dass sie sich Wissen zum Thema Evaluation aneignen mussten. Beide Prozentanteile liegen im Vergleich zu anderen abgefragten Wissensbereichen sehr weit oben (vgl. Hartz et al., 2007, S. 135-139).

Darüber hinaus konnte in einer Befragung von 166 Organisationen nachgewiesen werden, dass durch die LQW-Einführung besonders hohe Verbesserungen im Bereich Evaluation bezüglich des Wissens über die Teilnehmendenzufriedenheit erreicht werden konnten, während konkrete Lernerfolge vor und auch nach der LQW-Einführung weniger bekannt sind (vgl. Rädiker, 2006, S. 203). Der Vergleich mit anderen Qualitätsbereichen macht zudem deutlich, dass auch nach der Ersttestierung von einem weiteren Entwicklungsbedarf im Bereich der Evaluation auszugehen ist. Diese Ergebnisse werden von einer weiteren Studie zu den Auswirkungen von LQW bestätigt, die von 2006 bis 2010 unter knapp 250 Weiterbildungsorganisationen in

---

<sup>46</sup> Zum Zeitpunkt der Untersuchung lag das Modell in der 2. Version vor. Seit 2006 liegt es bis heute im Jahre 2012 unverändert in der 3. Version vor.

Deutschland und Österreich durchgeführt wurde (vgl. Rädiker, 2011). Die Verbesserungen im Bereich Evaluation bei den Ersttestierten fallen auch hier nur mittelmäßig aus und nach der ersten LQW-Testierung verbleibt der Evaluationsbereich im Vergleich zu den anderen Qualitätsbereichen im entwicklungsfähigen Mittelfeld. Im Zeitraum von der Anmeldung zur Restierung bis zur tatsächlichen Retestierung verbessert sich die Evaluation der Bildungsprozesse noch einmal und schließt deutlich zu den anderen Qualitätsbereichen auf. Sowohl vor als auch nach der Retestierung geben die Organisationen an, die Ideen und Verbesserungsvorschläge der Lehrenden zu berücksichtigen, weshalb sich diesbezüglich nur geringe Verbesserungen zeigen.

Neben diesen quantitativ angelegten Studien, die vornehmlich Aussagen über die Wirkungen im Bereich „Evaluation der Bildungsprozesse“ im Verhältnis zu den anderen Qualitätsbereichen ermöglichen, existieren auch Forschungsarbeiten zur konkreten Evaluationspraxis in Weiterbildungseinrichtungen, die das LQW-Modell anwenden. Busch (2009) untersuchte anhand von 15 Leitfadeninterviews, die auch einen Teil der Datenbasis für die vorliegende Arbeit darstellen, wie die befragten Organisationen die Evaluationsergebnisse verwenden. Es zeigte sich, dass die Nutzung der Evaluationsergebnisse sehr vielfältig ist und von der Nutzung zur Veränderung über die Nutzung zur vertiefenden Evaluation bis hin zur positiven Außendarstellung reicht, wobei sich interessanterweise keine Unterschiede zwischen Organisationen verschiedenen Typs oder verschiedener Größe identifizieren ließen. In einem weiteren Forschungsprojekt wurden Einzelfallanalysen der Evaluationspraxis in fünf unterschiedlichen LQW-testierten Bildungsorganisationen angefertigt und hierzu Fremdbeschreibungen der im LQW-Prozess entstandenen Selbstreporte angefertigt (vgl. Rädiker, 2010c). Die Analysen erlauben zwar keine Allgemeinaussagen über die Evaluationspraxis in LQW-Organisationen<sup>47</sup>, bieten jedoch einen intensiven Einblick in die innere Logik der Evaluationspraxis und ihrer Einbettung in die jeweilige Logik der Gesamtorganisation, die ebenfalls analysiert wurde (vgl. Zech, Dehn, Tödt, Rädiker, Mrugalla & Schunter, 2010). Auffälliges Ergebnis der Analysen war unter anderem, dass sowohl das jeweilige Evaluationsverständnis als auch die fokussierten Evaluationsgegenstände stark mit den jeweiligen Bildungsverständnissen korrelieren. So fokussiert eine Volkshochschule, deren habituellem Organisationstyp in der Analyse als „familiär“ bezeichnet wurde, gemäß ihrem Bildungsziel,

---

47 Zur sprachlichen Vereinfachung steht in dieser Arbeit der Begriff „LQW-Organisation“ für eine Organisation, die das LQW-Modell anwendet.



Zufriedenheit und persönliches Wohlbefinden der Teilnehmenden steigern zu wollen, auch bei der Evaluation die Zufriedenheit. Und ein kirchliches Bildungswerk, deren habituellem Typ als „dienende Organisation“ herausgearbeitet wurde, fokussiert getreu seines Bildungsverständnisses „Durch Bildung Aufgeschlossene in die Gesellschaft integrieren“ neben der Zufriedenheit auch die lernunterstützende Funktion von Evaluation (vgl. Rädiker, 2010c, S. 166).

Weitet man den Blick auf Forschungsergebnisse zur Evaluation in Weiterbildungsorganisationen jenseits von LQW so finden sich einige weitere interessante Ergebnisse, welche die Bedeutung der Evaluation in den Organisationen unterstreichen. Im Jahr 2008 haben das Bundesinstitut für Berufsbildung und das Deutsche Institut für Erwachsenenbildung den Weiterbildungsmonitor zum Thema „Personalentwicklung und wirtschaftliches Klima bei Weiterbildungsanbietern“ vorgelegt (vgl. Ambos, Egetenmeyer & Scheller, 2008). 45% der 1.720 befragten Weiterbildungsorganisationen geben an, einen eigenen Weiterbildungsbedarf im Bereich Qualitätsmanagement zu haben und 43% sehen einen Bedarf bezüglich der Evaluation der Bildungsarbeit, womit diese Bedarfe relativ weit oben auf den Plätzen 5 und 6 von 12 stehen. Diesen hohen Weiterbildungsbedarf in den Bereichen Qualitätsmanagement und Evaluation der Bildungsarbeit erklären die Autoren/innen der Studie mit den „zunehmenden Anforderungen an Evaluation und Zertifizierung (z.B. ISO, LQW) durch öffentliche Mittelgeber“ (ebd., S. 4). Sie weisen zudem darauf hin, dass der Weiterbildungsbedarf zu diesen Themen unter Organisationen, die sich mindestens zur Hälfte durch öffentliche Mittel finanzieren, noch höher liegt.

Zur konkreten Evaluationspraxis in Weiterbildungsorganisationen liegen nur wenige empirische Untersuchungen vor. Gutknecht-Gmeiner und Schlögl (2007) haben den Stand der Qualitätssicherung und -entwicklung bei österreichischen Institutionen der Erwachsenenbildung anhand einer Stichprobe von ca. 300 Organisationen analysiert. Sowohl als Maßnahme der Qualitätssicherung als auch als Maßnahme zur Sicherung pädagogischer und inhaltlicher Qualität nennen die Befragten am häufigsten Teilnehmendenbefragungen. Darüber hinaus kommt die Studie zu dem Schluss, „dass groß angelegte Evaluierungen zu Lernprozessen und -ergebnissen nicht regulärer Bestandteil institutioneller Qualitätssicherung sind“ (Gutknecht-Gmeiner, 2009, S. 13-8). Stattdessen ginge es vorwiegend um die Zufriedenheit der Teilnehmenden mit dem Lehr-Lern-Prozess auf der ersten Stufe nach Kirkpatrick (vgl. Abschnitt 2.6.1). Die zweite Kirkpatrick-Ebene zum Lernerfolg werde lediglich bei Zer-

tifikats-Angeboten berücksichtigt, die dritte und vierte Ebene zu Transfer und Wirkungen stellten ebenfalls Ausnahmefälle dar (vgl. Gutknecht-Gmeiner, 2009).

Auffällig ist, dass sich ein Großteil der Forschungsarbeiten im Feld der Weiterbildungsevaluation auf die berufliche Bildung mit besonderem Augenmerk auf die betriebliche Bildung richtet.<sup>48</sup> Häring (2003) geht beispielsweise mithilfe einer qualitativen Studie unter 22 Unternehmen den Fragen nach, welche Bedeutung Evaluation in diesen Organisationen zugeschrieben wird und welche Evaluationspraktiken hier existieren. Sie kommt dabei ebenfalls zu dem Ergebnis, dass sehr viele Unternehmen auf der Stufe der Zufriedenheit evaluieren, während nur wenige den Organisationserfolg einbeziehen (vgl. ebd., S. 257). Darüber hinaus stellt sich die Anwendung von Evaluationsinstrumenten in den Unternehmen sehr heterogen dar und wird im Wesentlichen durch die jeweilige Personalentwicklungskultur bestimmt (vgl. ebd., S. 254). Damit bestätigt sie unter anderem die von Wottawa (1996, S. 6) aufgestellte These, dass Evaluationsergebnisse nur dann sinnvoll verwertet werden können, wenn die Personalentwicklung eines Unternehmens einen so fortgeschrittenen Reifegrad aufweist, der Lernfähigkeit überhaupt erst ermöglicht. Häring legt zudem eine detaillierte Analyse weiterer Studien vor und kommt auf Basis der Resultate zu dem Schluss, „dass sich trotz erheblicher Fortschritte im wissenschaftlichen Bereich nur bedingt ein vermehrter Einsatz von Evaluierungsinstrumenten in der Praxis beobachten lässt“ (Häring, 2003, S. 76). Darüber hinaus haben laut der Studie die theoretischen Konzepte keinen Eingang in die Unternehmenspraxis gefunden und es besteht daher immer noch ein Mangel an Know-how und geeigneten Methoden.

Ein Teil der Forschungsanstrengungen im Bereich Weiterbildungsevaluation richtet sich schließlich darauf, geeignete Instrumente zu entwickeln oder bestehende zu optimieren. Dies geschieht vorwiegend, um eine hohe Aussagekraft der gewonnenen Daten zu erzielen. Häufig werden dabei aufwändige testtheoretische Verfahren angewendet, um valide und reliable quantitative Kursbeurteilungsbögen zu entwickeln (z.B. Hanke, 2005, für die Deutsche Bahn AG; Reischmann, 2006, für die allgemeine Weiterbildung). Zum Teil dient die Forschung aber auch dazu, neu entwickelte Instrumente hinsichtlich ihres Nutzens und ihrer Akzeptanz zu untersuchen (z.B. Buddenberg et al., 2007, zu Einsatzmöglichkeiten und Grenzen von Instrumenten der Lernerfolgskontrolle).

---

48 Veröffentlichungen zur Evaluation in der beruflichen Bildung betreffen z. B. die Messung von und Einflussfaktoren auf Weiterbildungserfolg (vgl. Bihler, 2006) oder verstehen sich als Sammelbände zu Theorie, Empirie und Praxis älteren und neueren Datums (vgl. Will et al., 1987b; Götz, 2001).

Vor dem Hintergrund der berichteten Studienergebnisse lässt sich zum Forschungsstand bezüglich der Evaluation von Bildungsprozessen in Weiterbildungsorganisationen folgendes Fazit ziehen: Im Gegensatz zur Lehrevaluation an Hochschulen, die man aufgrund der zahlreichen Publikationen<sup>49</sup> bereits als umfassend und facettenreich erforscht bezeichnen kann, finden sich bis auf wenige Ausnahmen, vor allem aus dem beruflichen Bereich, kaum empirische Daten über die konkrete Evaluationspraxis in Weiterbildungsorganisationen. Zugleich ist jedoch ein hoher Wissens- und Weiterbildungsbedarf bezüglich der Evaluation von Bildungsprozessen festzustellen. Worin dieser Wissens- und Weiterbildungsbedarf, man könnte sagen Kompetenzbedarf, genau besteht und vor welchen konkreten Problemen und Herausforderungen beispielsweise LQW-Organisationen bezüglich der Evaluation von Bildungsprozessen stehen, ist bisher jedoch unklar.

---

49 Um nur einige wenige Ausführungen zu Theorie und Praxis zu nennen: Balk (2000) befasst sich mit der Wirkung von Evaluationsrückmeldungen; Simonson und Pötschke (2006) haben die Akzeptanz internetgestützter Evaluation untersucht; Kromrey (2001b) beschäftigt sich vorwiegend auf konzeptioneller und methodischer Ebene mit Lehrevaluation; Rindermann (2003) legt Schlussfolgerungen zur Lehrevaluation aufgrund empirischer Ergebnisse vor.



## 4 Methodisches Vorgehen bei der durchgeführten Studie

Nachdem in den voranstehenden Kapiteln die Grundlagen für die durchgeführte Studie gelegt worden sind, liefert das vierte Kapitel eine detaillierte Beschreibung des methodischen Vorgehens bei der durchgeführten Studie. Hierzu werden im Abschnitt 4.1 zunächst die untersuchten Fragestellungen vorgestellt, woran sich eine detaillierte Darstellung der beiden verwendeten Datenquellen und der jeweiligen Auswertungstechniken anschließt (Abschnitte 4.2 und 4.3). Das Kapitel endet mit Hinweisen zum Umgang mit Quantifizierungen im Rahmen der qualitativen Analyse sowie zur Verallgemeinerbarkeit der Ergebnisse (Abschnitte 4.4 und 4.5).

### 4.1 Fragestellungen

Ausgehend von der in Abschnitt 3.3 ausgeführten Tatsache, dass es zwar Vermutungen und Mutmaßungen über die Evaluationspraxis in Weiterbildungsorganisationen gibt, aber nur wenige empirisch gestützte Informationen hierüber vorliegen, verfolgt die hier vorgestellte Studie einen Dreischritt, der sich in drei aufeinander aufbauenden Fragestellungen widerspiegelt.

In einem ersten Schritt wird der Status quo der Evaluationspraxis in Weiterbildungsorganisationen, die das LQW-Modell anwenden, beleuchtet:

*Fragestellung 1) Wie evaluieren LQW-Organisationen ihre Bildungsprozesse? Dabei geht es konkret um die Klärung der Detailfragen, zu welchem Zweck die Bildungsprozesse evaluiert, welche Daten mit welchen Instrumenten gewonnen, wie Daten ausgewertet und bewertet und wie Ergebnisse genutzt werden.*

Im zweiten Schritt wird der Blick auf die Probleme gerichtet, die in der Evaluationspraxis auftreten:

*Fragestellung 2) Vor welchen Herausforderungen stehen die Organisationen hinsichtlich der Evaluation ihrer Bildungsprozesse?*

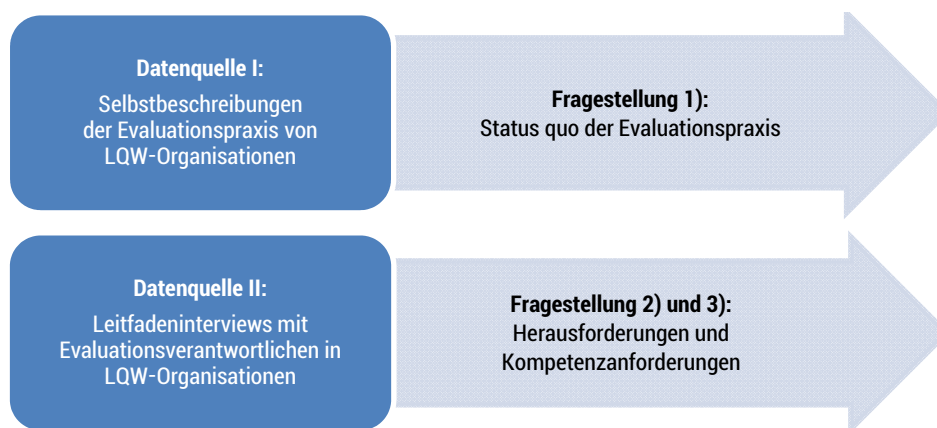
Vor dem Hintergrund der in Abschnitt 1.7 ausgeführten Debatte um Kompetenzen von Evaluierenden wird abschließend der dritten Fragestellung nachgegangen, die zum Ziel hat, die theoretischen Ausführungen und Studienergebnisse in einer Schlussfolgerung in Form eines Anforderungsprofil an Evaluierende in Weiterbildungsorganisationen zusammenzuführen:

*Fragestellung 3) Welche Kompetenzen benötigen Evaluierende in den Weiterbildungsorganisationen (unter Berücksichtigung der aktuellen Evaluationspraxis und der damit verbundenen Probleme), um Weiterbildungsprozesse erfolgreich zu evaluieren?*

Die Hauptkapitel, welche diesem Methodenkapitel folgen, greifen diesen Dreischritt auf. So werden in Kapitel 5 die Ergebnisse bezüglich des Status quo der Evaluationspraxis berichtet, Kapitel 6 informiert detailliert über die in der Evaluationspraxis auftretenden Probleme und Kapitel 7 ist den Evaluationskompetenzen gewidmet.

Die unterschiedlich gelagerten Fragestellungen legten nahe, nicht nur eine, sondern mehrere Datenquellen heranzuziehen. Zur Beantwortung der Fragestellungen wurde daher zum einen auf die Selbstbeschreibungen der Evaluationspraxis in 50 Selbstreporten von Organisationen, welche das LQW-Modell einsetzen, sowie zum anderen auf 24 leitfadengestützte Interviews mit Evaluationsverantwortlichen in LQW-Organisationen zurückgegriffen. Wie in Abb. 25 illustriert, dienten die Selbstbeschreibungen insbesondere dazu, die Fragestellung 1) zum Status quo der Evaluationspraxis zu beantworten, während die Leitfadeninterviews primär zur Beantwortung der Fragestellungen 2) und 3) zu den Herausforderungen in der Evaluationspraxis und den Kompetenzanforderungen beigetragen haben.

**Abb. 25: Zusammenhang zwischen Datenquellen und Forschungsfragen**



## 4.2 Selbstbeschreibungen der Evaluationspraxis (Datenquelle I)

Wie in Abschnitt 3.1 im Detail ausgeführt, sind Weiterbildungsorganisationen, die sich nach dem LQW-Modell testen lassen, verpflichtet, einen Selbstreport zu schreiben, in dem sie selbsterklärend und mit Nachweisen belegt darlegen, wie sie die Anforderungen von LQW in den elf Qualitätsbereichen erfüllen. Nach erfolgreicher Testierung liegt daher von jeder Organisation eine Selbstbeschreibung vor, die neben vielen anderen Aspekten auch detailliert Auskunft über das aktuelle Organi-

sationshandeln bezüglich der Evaluationspraxis gibt. Wie im vorigen Kapitel ausgeführt, umfasst dieser Teil der Selbstbeschreibung in der Regel drei bis fünf Seiten, ist häufig nach den Anforderungen gegliedert und informiert nicht nur über die aktuelle Situation, sondern beschreibt auch, „welche Änderungen die Organisationen in den letzten Jahren an ihrer Evaluationspraxis, wie den eingesetzten Evaluationsinstrumenten, dem Umfang oder dem Rhythmus, vorgenommen haben und aus welchen Gründen dies geschah“ (Rädiker, 2012, S. 38). Die Beschreibungen in den Selbstreporten haben den großen Vorteil, dass es sich um nicht-reaktive Daten<sup>50</sup> handelt, also um Daten, die aufgrund ihres Entstehungskontextes nicht durch den Forschungsprozess verzerrt worden sind. Es ist davon auszugehen, dass die Berichte über einen längeren Zeitraum entstanden sind, von mehreren Personen in der Organisation gelesen und korrigiert wurden und die Berichte von retestierten Organisationen sogar in der „2. überarbeiteten Auflage“ vorliegen, also einen Redigierungsprozess durchlaufen haben und nicht selten Auskunft über die Entwicklungen und Veränderungen der letzten vier Jahre geben. Da die Organisationen gezwungen sind, in ihren Selbstbeschreibungen die LQW-Anforderungen nicht einfach checklistenartig abzuheben, sondern verpflichtet sind, Handlungsweisen zu erläutern, zu begründen und auf ihr Leitbild und ihre Definition gelungenen Lernens zurückzubeziehen, können die Selbstbeschreibungen als Ausdruck der Organisationsidentität aufgefasst werden:<sup>51</sup>

„Die Anforderungen sind allerdings so offen formuliert, dass sie inhaltliche Ausführungen nicht präformieren. So wird z.B. gefordert, über Seminarevaluationen zu berichten. Es wird aber offen gelassen, welche Form, welchen Umfang und welchen Rhythmus die Evaluationen haben sollen. Oder es werden Führungsgrundsätze gefordert, ohne anzugeben, wie diese aussehen und was sie enthalten sollen. Alle Ausführungen in den Selbstreporten müssen sich durch Nachweise belegen lassen und von der Organisationsleitung legitimiert werden. Durch diesen Zwang zur Rückbindung an Nachweise und formale Entscheidungen der Weiterbildungsorganisationen können die Selbstreporte als Ausdruck der Identität und der Praxis der Gesamtorganisation behandelt werden.“ (Zech, 2010, S. 26)

---

<sup>50</sup> Vgl. Kromrey (2006, S. 532-534) zum Problem der Reaktivität von Datenerhebungsmethoden.

<sup>51</sup> Vgl. Zech et al. (2006, S. 74-80) zur Funktion von Selbstbeschreibungen und Fremdbeschreibungen im Kontext von LQW.

#### 4.2.1 Auswahl der Selbstreporte

Zum Zeitpunkt des Analysebeginns im Frühjahr 2009 lagen mehrere Hundert Selbstreporte von Organisationen in Deutschland vor. Weil davon auszugehen ist, dass sich viele Beschreibungen der Evaluationspraxis in den Selbstreporten ähneln, wurde entschieden, nicht mit dem gesamten, umfangreichen Datenmaterial zu arbeiten, sondern eine Stichprobe zu ziehen. Um eine geeignete Auswahl vorzunehmen, wurde zunächst eine Unterteilung der Organisationen in verschiedene Organisationstypen vorgenommen. Insgesamt wurden acht verschiedene Gruppen von Organisationen in der Grundgesamtheit<sup>52</sup> identifiziert:

1. sozial und/oder kulturell orientierte Organisationen
2. weltanschaulich gebundene Organisationen
3. Volkshochschulen
4. kommerziell orientierte Organisationen
5. wirtschaftsnahe Organisationen
6. Landesverbände z.B. von Volkshochschulen
7. Weiterbildungsabteilungen von Unternehmen
8. universitäre bzw. wissenschaftsnahe Organisationen

Da zu den letzten drei Gruppen jeweils nur sehr wenige Organisationen zählen und es beim Einbezug dieser Organisationen in die Analyse unter anderem Probleme bei der Wahrung der Anonymität hätte geben können, wurden sie nicht berücksichtigt. Da zudem auch nur wenige wirtschaftsnahe Organisationen LQW anwenden, wurden sie im Analyseprozess mit der Gruppe der kommerziell orientierten Organisationen zusammengelegt, sodass schlussendlich in der Stichprobe die folgenden vier verschiedenen Organisationstypen vertreten waren:

**A) sozial und/oder kulturell orientierte Organisationen** In diese Gruppe fallen alle Organisationen, deren Bildungsarbeit sich entweder an Personen richtet, die im sozialen Feld arbeiten (z.B. Erzieher/innen im Kindergarten oder in Behinderteneinrichtungen) oder die selbst sozial benachteiligt sind (z.B. arbeitslose Frauen oder Migrantinnen). In Abgrenzung zur vierten Gruppe (kommerziell/wirtschaftsnahe Organisationen) überwiegt bei diesen Organisationen der soziale Aspekt die berufli-

---

<sup>52</sup> Eine Auflistung der aktuell im LQW-Verfahren befindlichen Organisationen findet sich unter [www.artset-lqw.de](http://www.artset-lqw.de), Menüpunkt „Organisationen“.



che Bildungsarbeit. Weiterhin umfasst diese Gruppe Heimvolkshochschulen<sup>53</sup> und Träger der politisch freien und ungebundenen Bildungsarbeit, die beide vorwiegend mehrtägige Seminare, in der Regel auch mit Übernachtungsmöglichkeiten und Verpflegung, anbieten. Zur sprachlichen Vereinfachung wird diese Gruppe im Folgenden teilweise auch mit „sozial/kulturell orientierte Organisationen“ betitelt, wenngleich einzelne Organisationen nur dem sozialen oder dem kulturellen Feld zugeordnet werden können.

**B) weltanschaulich gebundene Organisationen** Zu dieser Gruppe gehören Organisationen von Parteien, Gewerkschaften und Kirchen, die unter anderem Weiterbildungen für Betriebsräte, Schwerbehindertenvertretungen, Sozialpädagogen/innen und Mitarbeitende der Kirchen anbieten. Allen Organisationen ist gemeinsam, dass sie sich in einer Trägerschaft befinden, die (implizit) zu bestimmten Werten und Weltanschauungen verpflichtet. Größtenteils bieten die Organisationen mehrtägige Seminare im eigenen Bildungshaus an.

**C) Volkshochschulen** Diese Gruppe umfasst sowohl städtische Volkshochschulen als auch Volkshochschulen von Landkreisen, welche in der Regel ein breites Spektrum an Angeboten der allgemeinen Erwachsenenbildung offerieren. Volkshochschulverbände der Länder, die teilweise auch LQW testiert sind, fallen jedoch nicht in diese Kategorie. Als klassische Veranstaltungsform der Volkshochschule gilt der wöchentlich stattfindende Kurs, doch sind auch viele geblockte Veranstaltungen, wie etwa Tages- oder Wochenendkurse, im Kursprogramm der zahlreichen Volkshochschulen zu finden.

**D) kommerziell orientierte und/oder wirtschaftsnahe Organisationen** Zu dieser Gruppe zählen alle Organisationen, die dem Profit-Bereich zuzuordnen sind und die fast ausnahmslos berufliche Bildung anbieten. In diese Kategorie fallen damit sowohl Sprach- und Computerschulen wie auch Organisationen, die „klassische Arbeitsagentur-Kurse“, von der Ausbildung über die Umschulung bis zur Integrationsmaßnahme, durchführen. Zu den wirtschaftsnahen Organisationen zählen vorwiegend die Bildungswerke der Wirtschaft. Die Veranstaltungsformen der kommerziell orientierten, wirtschaftsnahen Organisationen sind sehr heterogen und decken das ganze denkbare Spektrum von wöchentlichen Abendkursen bis hin zu mehrmonatigen Kursen ab. Zur sprachlichen Vereinfachung wird diese Gruppe im Folgenden

---

53 Anders als ihr Name suggerieren mag, zählen Heimvolkshochschulen aufgrund ihres meist engeren Bildungsangebots, ihrer kompakteren Veranstaltungsformen und der häufig in Bildungsmaßnahmen integrierten Übernachtungsmöglichkeiten nicht zur Gruppe der Volkshochschulen.

teilweise auch mit „kommerziell orientierte/wirtschaftsnahe Organisationen“ betitelt, wenngleich nicht alle kommerziell orientierten Organisationen zugleich als wirtschaftsnah zu charakterisieren sind.

Im April 2009 wurden 50 Selbstreporte von Weiterbildungsorganisationen in Deutschland, die sich zur ersten Testierung oder ersten Retestierung angemeldet hatten, per Zufall ausgewählt. Für die Analyse wurden nur die Ausführungen zum Qualitätsbereich 5 „Evaluation der Bildungsprozesse“ verwendet. Zu jeder Selbstbeschreibung wurde neben dem Organisationstyp auch die Organisationsgröße festgehalten. Analog zu anderen Forschungsstudien über LQW-Organisationen wurde dabei zwischen kleinen Organisationen mit weniger als 5 hauptamtlichen pädagogischen Mitarbeitenden (HPM), mittleren Organisationen mit 5 bis 15 HPM sowie großen Organisationen, die mehr als 15 HPM beschäftigen, unterschieden (vgl. Rädiker, 2006, 2011). Dieser Einteilung liegt unter anderem die Überlegung zugrunde, dass eine Organisation mit mehr als vier hauptamtlichen pädagogischen Mitarbeitenden (und die teilweise hinzukommenden Leitungs- und Verwaltungskräfte) nicht mehr „auf Zuruf“ gesteuert werden kann.<sup>54</sup> Abb. 26 zeigt, wie sich die Organisationen auf die drei Merkmale Typ, Größe und Retestierung verteilen.

**Abb. 26: Zusammensetzung der Zufallsstichprobe (n = 50) nach Organisationstyp, Organisationsgröße und Retestierung**

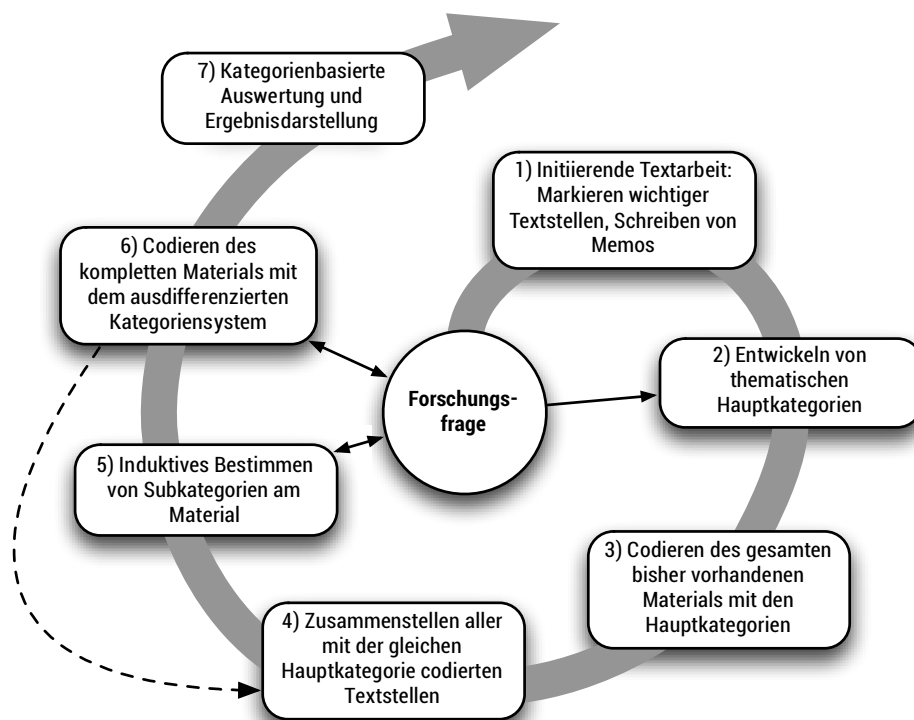
	absoluter Anteil	relativer Anteil
<b>ORGANISATIONSTYP</b>		
sozial/kulturell orientiert	10	20%
weltanschaulich gebunden	9	18%
Volkshochschulen	20	40%
kommerziell orientiert/wirtschaftsnah	11	22%
<b>ORGANISATIONSGRÖSSE</b>		
klein (weniger als 5 HPM)	17	34%
mittel (5–15 HPM)	22	44%
groß (mehr als 15 HPM)	11	22%
<b>RETESTIERUNG</b>		
ja	43	86%
nein (= Ersttestierung)	7	14%

<sup>54</sup> Außen vor bleibt bei dieser Einteilung die Anzahl der freiberuflichen Lehrenden, die von Organisation zu Organisation stark variieren kann. Da jedoch davon auszugehen ist, dass die Evaluation größtenteils von den hauptamtlichen pädagogischen Mitarbeitenden konzipiert und organisiert wird, erscheint die Vernachlässigung der freiberuflichen Lehrenden vertretbar.

#### 4.2.2 Die qualitative Inhaltsanalyse

Die 50 Selbstbeschreibungen der Organisationen über ihre Evaluationspraxis wurden einer inhaltlich strukturierenden qualitativen Inhaltsanalyse nach Kuckartz (2012) unterzogen.<sup>55</sup> Dabei handelt es sich um eine Form der kategorienbasierten Auswertung qualitativer Daten. Kuckartz unterscheidet bei diesem Analyseverfahren ursprünglich sieben Phasen, die in der vereinfachten Darstellung des Vorgehens bei dieser Studie in Abb. 27 zu sechs Phasen verdichtet wurden.

Abb. 27: Schematischer Ablauf der inhaltlich strukturierenden qualitativen Inhaltsanalyse (Quelle: Kuckartz, 2012, S. 78)



##### *Phase 1: Initiierende Textarbeit*

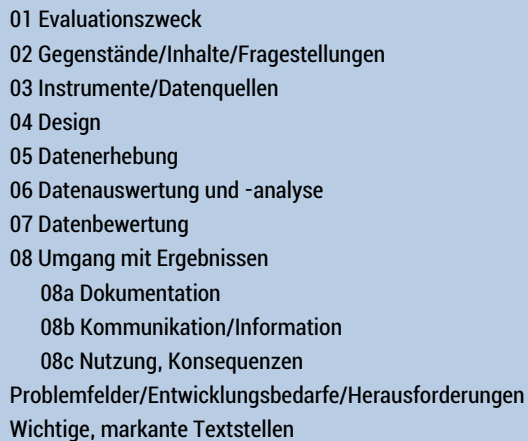
Um sich mit dem vorliegenden Datenmaterial vertraut zu machen, wurden die Selbstbeschreibungen zunächst gelesen, wobei Auffälligkeiten mit Blick auf die Forschungsfragen vermerkt wurden. Im Anschluss wurde als weitere Annäherung an das Material und als erster Analyseschritt zu jeweils einem der vier Organisationstypen eine verdichtende Fallzusammenfassung, ein sogenanntes „Case Summary“ (vgl. Kuckartz et al., 2008, S. 33-34; 2009, S. 67-72; Kuckartz, 2012) erstellt. Die erstellten Case Summaries folgen in ihrem Aufbau einem typischen Evaluationsablauf

<sup>55</sup> Die Inhaltsanalyse wurde mithilfe der Analysesoftware MAXQDA vorgenommen ([www.maxqda.de](http://www.maxqda.de)). Grundideen dieser Analysetechnik wie etwa die Grob- und Feincodierung und ihrer Unterstützung durch Computersoftware sind auch bei Kuckartz und Rädiker (2010a, 2010b) zu finden.

und entsprechen somit nicht bloß einer sequenziell verkürzten Darstellung des Textes, sondern führten bereits eine für die spätere Analyse hilfreiche Ordnung und Systematik ein. Die Case Summaries enthalten keine Wertungen oder Interpretationen, sondern geben nur den vorhandenen Inhalt in kondensierter und gegliederter Form wieder. Die Summaries wurden zunächst stichpunktartig verfasst und in einem weiteren Schritt ausformuliert. In eckigen Klammern wurden dabei Interpretationen, Auffälligkeiten und Notizen festgehalten, wie z.B. beim ersten Case Summary „[Die Evaluation dient also im Wesentlichen der Erfolgskontrolle. Zufriedenheit wird hier auf spezielle Bereiche bezogen und nicht global abgefragt.]“ oder beim zweiten „[Die Organisation benennt im Vergleich zu anderen sehr viele und detaillierte Evaluationszwecke.]“. Diese Hinweise gehen über die rein beschreibende Ebene hinaus und waren für die anschließende Analyse von Bedeutung, etwa für die Bildung geeigneter Kategorien.

#### *Phase 2: Entwicklung thematischer Hauptkategorien*

Im nächsten Schritt wurden für die kategorienbasierte Auswertung thematische Hauptkategorien gebildet, die sich aus verschiedenen Quellen speisten: aus der Fragestellung zum Status quo, aus den Überlegungen in Kapitel 1 und Kapitel 2, aus früher durchgeführten Analysen zur Evaluationspraxis (ebenfalls am Datenmaterial „Selbstreport“, Rädiker, 2010c) sowie aus den Anforderungen, die das LQW-Modell im Qualitätsbereich „Evaluation der Bildungsprozesse“ vorsieht. Wie in Abb. 28 erkennbar, folgen die Hauptkategorien in ihrer Logik einem typischen Evaluationsablauf, wie er in Abschnitt 2.6.4 beschrieben wurde, und sind entsprechend nummeriert. Für Textstellen mit hervorstechenden oder besonders aussagekräftigen Inhalten wurde zusätzlich eine Kategorie „Wichtige, markante Textstellen“ aufgenommen. Im Verlauf der anschließenden Phase 3 wurde das Kategoriensystem an das empirische Material angepasst und geringfügig überarbeitet; so stellte sich heraus, dass viele Organisationen in den Selbstbeschreibungen auch Probleme, Herausforderungen und Entwicklungsbedarfe berichten, die für die zweite Fragestellung relevant sind, und dass die ursprünglich aus einem typischen Evaluationsablauf abgeleitete Kategorie „Schlussfolgerungen ziehen“ keine Entsprechung im empirischen Material fand. Zu Beginn der Analyse wurde für jede Kategorie eine Definition formuliert, die festlegt, welche Inhalte mit der Kategorie erfasst werden. Die Kategoriendefinitionen, die im Laufe der Analyse fortgeschrieben, überarbeitet und um Ankerbeispiele (Originalzitate) bereichert wurden, sind im Anhang A) aufgeführt.

**Abb. 28: Hauptkategorien nach Abschluss der Analyse**


01 Evaluationszweck  
 02 Gegenstände/Inhalte/Fragestellungen  
 03 Instrumente/Datenquellen  
 04 Design  
 05 Datenerhebung  
 06 Datenauswertung und -analyse  
 07 Datenbewertung  
 08 Umgang mit Ergebnissen  
     08a Dokumentation  
     08b Kommunikation/Information  
     08c Nutzung, Konsequenzen  
 Problemfelder/Entwicklungsbedarfe/Herausforderungen  
 Wichtige, markante Textstellen

*Phase 3: Codierung mit den Hauptkategorien*

In der dritten Phase wurden in allen 50 Selbstreporten die Textsegmente mit für die Fragestellung relevanten Inhalten den in Phase 2 entwickelten Hauptkategorien zugeordnet. Zur Arbeitsunterstützung, aber auch zur Qualitätssteigerung der vorgenommenen Codierungen, wurde diese Phase der Grobcodierung zusammen mit der Unterstützung zweier studentischer Hilfskräfte bewältigt, die mit dem Themengebiet vertraut waren (vgl. Abschnitt 4.3.2 zur Steigerung der Auswertungsqualität). Dem Codierprozess lagen in Anlehnung an Kuckartz et al. (2008, S. 39-40) folgende Regeln zugrunde:

1. Es werden Sinneinheiten codiert, mindestens ein Satz bzw. Teilsatz. Wichtig ist, dass die codierte Stelle auch ohne den sie umgebenden Text verständlich bleibt.
2. Wenn eine Information in einem Selbstreport zweimal enthalten ist, wird sie nur einmal codiert, um Redundanzen bei der späteren verdichtenden Auswertung zu vermeiden. Dabei wird tabellarischen Übersichten und Zusammenfassungen der Vorzug gegeben, weil aus ihnen auf einen Blick mehr Informationen entnehmbar sind und man meist nur eine statt mehrerer Codierungen benötigt.

*Phasen 4, 5 & 6: Entwicklung von Subkategorien am Material & Codierung mit den Subkategorien*

Für jede Hauptkategorie wurde im Anschluss an die Grobcodierung der folgende Prozess, teilweise in Iterationsschleifen, durchlaufen:<sup>56</sup>

---

<sup>56</sup> Früh (2001) beschreibt diesen Prozess als sogenannte „Basiswissengeleitete offene Kategorienfindung“ in einem Buchbeitrag gleichen Namens. Er unterteilt den Prozess in drei grobe Schritte, die den hier vorgestellten sehr ähnlich sind: 1. Selektion/Reduktion (Auswahl von Textstellen mit Relevanz für Forschungsfrage. Redundante und inhaltsleere Stellen werden ausgeblendet.), 2. Bündel-

- a) Zusammenstellung der codierten Textstellen einer Hauptkategorie,
- b) Durchlesen von etwa fünf bis zehn Textstellen, um sich einen ersten Überblick über die Inhalte und die angesprochenen Dimensionen zu verschaffen,
- c) Bildung von Subkategorien für jeden neuen inhaltlichen Aspekt (beginnend mit der ersten Textstelle),
- d) Zuordnung von Textstellen – in der Regel Ausschnitte aus den Grobcodierungen – wenn möglich zu vorhandenen Subkategorien, ansonsten Neubildung von Subkategorien,
- e) währenddessen Anpassung und Erweiterung der Subkategorien durch Gruppierung und Ausdifferenzierung sowie durch Präzisierung der Kategorientitel.

Zum Teil wurden die neu erstellen Kategorien direkt in der Analysesoftware angelegt, zum Teil wurden die Textstellen einer Hauptkategorie zunächst zu Excel exportiert, um mit einer Zwei- bzw. Drei-Spalten-Ansicht arbeiten zu können: Die erste Spalte enthielt pro Zeile eine Textstelle der Hauptkategorie, in der zweiten Spalte wurden die jeweilige Reduktionen und Abstraktionen der Textstellen notiert. Die Abstraktionen wurden mit Blick auf die bereits vorgenommenen Abstraktionen formuliert, so dass sich Ähnlichkeiten herausbildeten. Teilweise wurden vorherige Abstraktionen angepasst. Wenn die Abstraktionen und Reduktionen nicht bereits als Kategorientitel dienen konnten, wurden passende Kategorientitel in einer dritten Spalte festgehalten. Der große Vorteil der mehrspaltigen Tabellendarstellung lag darin, dass auf einen Blick alle Original-Textstellen und die ihr zugeordneten Subkategorien sichtbar waren und leicht Anpassungen vorgenommen werden konnten.

Bei der Bildung der Subkategorien zu einer Hauptkategorie wurde darauf geachtet, die neu gebildeten Kategorien möglichst auf dem gleichen Abstraktionsniveau zu halten. Die Generierung der Subkategorien orientierte sich im Wesentlichen am Vorwissen zum Forschungsgegenstand, das in den vorhergehenden Kapiteln ausführlich dargelegt wurde. Als Orientierungspunkt und Ideenpool dienten beispielsweise die theoretischen Überlegungen zu den vier Dimensionen Lernen, Lehren, Lerninfrastruktur und Lernorganisation, die alle als Subkategorien verwendet wurden, um die zahlreichen Evaluationsinhalte und -gegenstände zu fassen. Als hilfreich für die Kategorienbildung, speziell für die Ordnung der Kategorien, stellten sich auch

---

lung (Gruppierung nach inhaltlichen Gemeinsamkeiten auf einheitlicher Abstraktionsebene), 3. Generalisierung/Abstraktion (jedes Bündel erhält ein Label, also einen Kategorientitel.)

die LQW-Spezifikationen heraus, die für jeden Qualitätsbereich formuliert sind. So konnten bei der Frage, ob die Lernmaterialien besser als Unterkategorie des „Lernens“ oder der „Lerninfrastruktur“ aufgefasst werden, die Spezifikationen hierauf Antwort geben, denn dort werden die Lernmaterialien zur Lerninfrastruktur gezählt. Abb. 29 zeigt exemplarisch einen Auszug aus den Subkategorien der Hauptkategorie „Gegenstände/Inhalte/Fragestellungen“ und illustriert, dass auch unterhalb der Subkategorien weitere Unterkategorien aus Systematisierungsgründen angelegt wurden.

**Abb. 29: Auszug der Subkategorien zur Hauptkategorie „Gegenstände/Inhalte/Fragestellungen“**

02 Gegenstände/Inhalte/Fragestellungen
Lernorganisation
Service/Mitarbeiter
...
Lerninfrastruktur
Infrastruktur allgemein
...
Lehren
Dozent/Unterricht/Methoden
...
Lernen
Lernerfolg
...
Zufriedenheit
overall/nicht weiter spezifiziert
...
Definition gelungenen Lernens/Leitbild
...
Sonstiges

Wie in der letzten Zeile von Abb. 29 exemplarisch für die Kategorie „02 Gegenstände/Inhalte/Fragestellungen“ zu sehen, erhielt jede Hauptkategorie eine Subkategorie „Sonstiges“, die als Sammelbecken für Nennungen, die nur bei wenigen (meist unter fünf) Organisationen vorkamen, und für pauschalisierende, sehr allgemeine Aussagen ohne Erkenntnisgewinn genutzt wurde.

Nach Codierung aller Textstellen einer Hauptkategorie mit den neu erstellten Subkategorien wurden die Inhalte jeder Subkategorie auf interne Konsistenz geprüft. Bei schwierigen Abgrenzungsproblemen wurden Kategoriendefinitionen ergänzt und gegebenenfalls auch Subkategorien weiter ausdifferenziert oder adaptiert. Sowohl bei der Überprüfung der Zuordnungen als auch bei der Bildung der Kategorien hat sich im Analyseprozess das Arbeiten mit kurzen Inhaltsverdichtungen der einzelnen codierten Textstellen als sehr hilfreich herausgestellt: Mit der Kommentarfunktion

der Auswertungssoftware MAXQDA wurden zu zahlreichen codierten Textstellen Zusammenfassungen erstellt, um einen schnellen Überblick über die Inhalte einer Subkategorie erhalten zu können (vgl. Abb. 30). Dieser Überblick erlaubte zum einen unpassende Textstellen in passende Subkategorien zu verschieben, zum anderen aber auch Subkategorien mit zahlreichen zugeordneten Textstellen weiter zu differenzieren.

**Abb. 30: Beispiel für einige Zusammenfassungen von Textstellen der Subkategorie „Erfolgskontrolle, Zielüberprüfung“ der Hauptkategorie „Evaluationszweck“**

▲ Kommentar
Überprüfung der Erfüllung der Ansprüche aus Leitbild - konkret beschrieben
Überprüfung der Erfüllung von Qualitätsansprüchen
Überprüfung der Realisierung der DgL
Überprüfung der selbst formulierten Ansprüche
Überprüfung der Ziele aus dem Leitbild und der DgL
Überprüfung des eigenen Anspruchs hinsichtlich Kursinhalt, Dozententätigkeit etc.

#### *Phase 7: Kategorienbasierte Auswertung und Ergebnisdarstellung*

Im Anschluss an die Codierarbeit wurden die Ergebnisse verschriftlicht und aufbereitet, indem die identifizierten Inhalte zur Evaluationspraxis zu den einzelnen Hauptkategorien berichtet und teilweise grafisch illustriert wurden. Zur Auswertung wurden unter anderem auch Kreuztabellen angefertigt, um Unterschiede zwischen verschiedenen Organisationstypen identifizieren zu können. Bei der Darstellung der Ergebnisse wurden darüber hinaus einzelne Themen miteinander in Beziehung gesetzt.

#### **4.2.3 Maßnahmen zur Steigerung der Auswertungsqualität**

Wenngleich die Bildung von Kategorien häufig als „Kunstlehre“ beschrieben wird und es sicher unbestreitbar ist, dass dabei ein nicht vermittelbarer Akt der Kreativität beteiligt ist, so darf nicht der Eindruck entstehen, Kategorienbildung und eine darauf aufbauende Analyse sei generell willkürlich und verwehre sich jeder Qualitätsüberprüfung und Erhöhung ihrer Zuverlässigkeit. Im Gegenteil: Es lassen sich zahlreiche Verfahren und Strategien anwenden, welche die Qualität der computer-gestützten qualitativen Inhaltsanalyse erhöhen und welche in diesem Forschungsprojekt auch zum Einsatz kamen. Anders als in der quantitativen Sozialforschung, in der standardisierte Verfahren den Alltag beherrschen und breit akzeptierte Gütekriterien wie Objektivität, Reliabilität, Validität und hierzu teilweise berechenbare Koeffizienten existieren, wird in der qualitativen Sozialforschung für Verfahrensdoku-

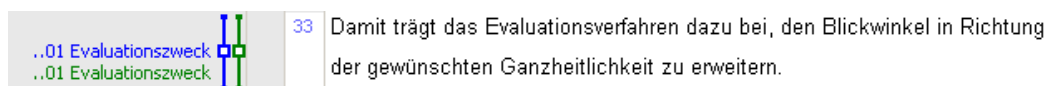


mentation und Transparenz, für Regelgeleitetheit und Argumentation plädiert (vgl. zur Diskussion um Qualität und Gütekriterien in der qualitativen Sozialforschung Flick, 2007; Mayring, 2002; Grunenberg, 2001; Steinke, 2007). In der durchgeführten Studie wurden folgende Strategien und Verfahrensweisen angewendet, um die Qualität des Codierprozesses zu sichern und zu erhöhen:

#### *Überprüfung der InterCoder-Übereinstimmung*

Als zentrales Mittel der Qualitätsverbesserung diente die Überprüfung der Übereinstimmung unterschiedlicher Codierer/innen. Zu diesem Zweck wurde das Codesystem verdoppelt und für zwei Coder unterschiedlich eingefärbt (vgl. Rädiker & Stefer, 2007, S. 86-87). Die ersten fünf Texte wurden von einer studentischen Hilfskraft und mir unabhängig voneinander codiert und anschließend in einer Projektdatei vereint, in der (Nicht-)Übereinstimmungen durch die unterschiedlichen Farben direkt sichtbar waren, wie Abb. 31 zeigt.

**Abb. 31: Überprüfung der InterCoder-Übereinstimmung mithilfe unterschiedlicher Farben**



Obwohl die Berechnung von Koeffizienten der InterCoder-Übereinstimmung nicht nur in der quantitativen, sondern auch in der qualitativen Sozialforschung Anwendung findet (vgl. Rädiker, 2010a), wurde in dieser Studie einem konsensuellen, kommunikativen Überprüfungsprozess der Vorzug gegeben, in dem die Nicht-Übereinstimmungen besprochen und die der Zuordnung zugrunde liegenden Probleme reflektiert und geklärt wurden (vgl. Hopf & Schmidt, 1993, S. 61-63). Der Vergleich der Codierungen für die ersten fünf Texte zeigte sehr viele Übereinstimmungen, lediglich an wenigen Stellen war die Überschrift oder ein zusätzlicher Satz mit- bzw. nicht mitcodiert worden oder es befanden sich mehrere Codierungen mit dem gleichen Code dicht gefolgt hintereinander, statt dass die Textstelle mit einer durchgängigen Codierung versehen war.

Als problematisch stellte sich jedoch die genaue Abgrenzung der Kategorien „01 Evaluationszweck“ und „08c Nutzung“ heraus. Zur Schärfung der Trennlinie zwischen den Kategorien wurden die Definitionen beider Kategorien angepasst und erweitert: Solange es sich um Absichtserklärungen handelt, was mit der Evaluation erreicht werden soll, was sich z.B. in Formulierungen wie „Evaluation soll“ oder „mit der Evaluation wollen wir“ zeigt, wurde die entsprechende Textstelle mit „Evaluationszweck“ codiert. Bei Aussagen, die über die Absichtserklärung hinausgehen, wur-

de der Code „Nutzung“ verwendet. Der Kategorientitel „Nutzung“ wurde zudem um den Begriff „Konsequenzen“ erweitert, weil mit dieser Kategorie explizit Aussagen erfasst werden sollten, die über den Umgang mit Ergebnissen im Sinne von Konsequenzen informieren. Neben dieser Abgrenzungsproblematik stellte sich zudem auch der Umgang mit Doppelcodierungen als schwierig heraus, da nicht immer klar war, ab wann eine Information als doppelt anzusehen ist. Da es jedoch inzwischen möglich geworden war, mit der weiterentwickelten Analysesoftware Auswertungen vorzunehmen, die jeden Code pro Text nur einmal werten, wurde diese Unklarheit als weniger problematisch eingestuft. Als Codieranweisung wurde festgehalten, dass Doppelcodierungen vermieden, inhaltliche Doppelungen am Text mit einer Notiz in Form eines sogenannten Memos kenntlich gemacht und im Zweifelsfall lieber doppelt codiert werden sollten. Zudem sollten bei inhaltlichen Doppelungen, den Informationen in Tabellen der Vorzug gegeben werden, weil aus ihnen auf einen Blick mehr Informationen entnehmbar sind und man meist nur eine statt mehrerer Codierungen benötigt.

Zur weiteren Qualitätsverbesserung hat eine zweite studentische Hilfskraft die Codierung der Texte mit den Hauptkategorien unterstützt, so dass beide Codiererinnen mit den neuen Anweisungen und Kategoriendefinitionen weitere fünf Texte durchgearbeitet haben und wir anschließend zu dritt eine konsensuelle Klärung der sehr wenigen Problemstellen herbeigeführt haben. Als Resultat unserer Diskussion ergänzten wir als Codieranweisung den Hinweis, nicht zu viel umgebenden Text bei einer relevanten Stelle zu codieren, weil ansonsten nicht mehr auf den ersten Blick erkennbar sein kann, auf welchen Inhalt sich die Codierung bezieht. Zudem wurde vorgeschrieben, vor dem Codieren die Struktur des Textes anzuschauen, um frühzeitig Hinweise auf Redundanzen zu erhalten. Die weiteren vierzig Texte wurden allein von den beiden studentischen Hilfskräften mit den Hauptkategorien codiert und im Anschluss von mir überprüft, in dem ich Hauptkategorie für Hauptkategorie die codierten Textstellen kontrolliert habe.

In allen Phasen des gemeinsamen Codierprozesses zu den Hauptkategorien wurde ein zusätzlicher Code mit dem Titel „Zuordnungsprobleme“ aufgenommen, in dem die beiden studentischen Hilfskräfte Textpassagen ablegen konnten, die sie als relevant für die Forschungsfrage eingeschätzt hatten, jedoch nicht zweifelsfrei zuordnen konnten. Von dieser Möglichkeit haben die beiden Codiererinnen bei weniger als zehn Textstellen Gebrauch gemacht, was als Bestätigung der Qualität der Kategoriendefinitionen gesehen wurde. Während des Intercoder-Prozesses haben wir ge-

meinsam, nach Abschluss dieses Prozesses habe nur ich allein die endgültige Zuordnung der fraglichen Textstellen vorgenommen.

#### *Überprüfung der Vollständigkeit*

Nach Abschluss der Codierung mit den Hauptkategorien wurde eine Vollständigkeitskontrolle mit Hilfe des Code-Matrix-Browsers in MAXQDA vorgenommen, einer tabellenartigen Visualisierung, welche in den Zeilen alle Codes und in den Spalten alle Texte listet und in jeder Zelle mit grafischen Symbolen anzeigt, ob ein Code in einem Text vergeben wurde (vgl. Kuckartz, 2010a, S. 193-194). Texte, in denen eine Kategorie überhaupt nicht zugeordnet war, wurden nach dem entsprechenden Thema durchsucht und falls noch passende (zuvor übersehene) Textstellen identifiziert werden konnten, erfolgte eine Codierung der Textstelle. Die Kontrolle auf Vollständigkeit wurde jedoch nicht nur mit dem Code-Matrix-Browser, sondern auch mit Suchfunktionen realisiert. Beispielsweise war zunächst nur in 31 Dokumenten „Zufriedenheit“ mit „02 Gegenstand, Inhalt, Fragestellung“ codiert. Ein lexikalischer Suchlauf zeigte jedoch, dass der Begriff „zufrieden“ in 42 Dokumenten vorkommt, woraufhin die elf zusätzlichen Dokumente noch einmal durchgeschaut wurden und in sieben Texten fehlende Codierungen ergänzt wurden. Enthielt ein Selbstreport keine Angaben zu einer Kategorie, wurde dies in einer Notiz am Textanfang vermerkt.

#### *Bildung von Subkategorien im Team*

Während die Hauptkategorien vorwiegend aus theoretischen Überlegungen abgeleitet wurden, orientierte sich die Entwicklung der Subkategorien größtenteils am vorliegenden empirischen Material. Um in diesem Prozess ebenfalls eine hohe Qualität zu erreichen, wurde teilweise mit einer weiteren studentischen Hilfskraft im Zweier-Team gearbeitet, und zwar mit zwei verschiedenen Vorgehensweisen: Bei einigen Kategorien bildeten wir unabhängig voneinander Kategorien, verglichen gemeinsam unsere Ergebnisse und erarbeiten ein endgültiges System an Subkategorien; bei anderen Kategorien erarbeiteten wir diskursiv ein System ohne vorherige alleinige Schritte.

#### *Schreiben von knappen Zusammenfassungen zu Textstellen*

Wie weiter oben zur Kategorienbildung am Material bereits ausgeführt und wie in Abb. 30 dargestellt, wurden zu fast allen codierten Textstellen knappe Zusammenfassungen geschrieben, welche eine optimale Übersicht und schnelle Durchsicht über die Inhalte einer Kategorie und eine einfache Prüfung der internen Konsistenz

von Subkategorien erlaubt. Bei einigen Kategorien mit zahlreichen Textstellen und einem komplexen System an Subkategorien wurden während des Prozesses der Kategorienbildung und ersten Zuordnung von Textstellen die Kommentare von Textstellen mit dem Hinweis „OK“ versehen, um deutlich zu machen, wenn es sich um eine endgültig entschiedene Codierung handelt.








#### *Überprüfung der Sonstiges-Kategorie*

Innerhalb jeder Hauptkategorie wurde eine Kategorie „Sonstiges“ definiert, in der Einzelnennungen und zunächst nicht zuordnbare Textstellen gesammelt und zumeist mit einer kurzen Inhaltszusammenfassung versehen wurden. Nach Abschluss der Entwicklung von Subkategorien und der Zuordnung der Textstellen zu den Subkategorien wurden alle Einträge der Sonstiges-Kategorie überprüft, inwieweit sie entweder a) eine eigene neue Subkategorie bilden, weil sie sich zu einem bisher nicht bedachten Aspekt gruppieren lassen, oder b) bereits vorhandenen Kategorien zuordnen lassen.

#### *Verwendung unterschiedlicher Memosymbole*

Das Analyseprogramm MAXQDA erlaubt es, an zahlreiche Stellen (z.B. an einem Gesamttext, an einer Textstelle oder zu einer Kategorie) sogenannte Memos mit Notizen zu heften. Um Memos verschiedener Zwecke voneinander zu unterscheiden, kamen – wie in Abb. 32 gezeigt – verschiedene Memosymbole zum Einsatz.

**Abb. 32: Verschiedene Memosymbole für Notizen mit unterschiedlicher Funktion**

	1	Allgemeine Memos im Text, Dokument-Memos und Code-Memos
	2	Text ist gut geeignet für CaseSummary
	3	CaseSummary
	4	Hinweise auf Dopplungen im Text
	5	zu klärende Punkte
	6	Auswertung, Theoriehinweise, Erkenntnisse
	7	Hinweise, wenn Themen trotz intensiver Suche nicht im Text vorkommen

### **4.3 Leitfadeninterviews mit Evaluationsverantwortlichen (Datenquelle II)**

Zwar finden sich auch in den Selbstreporten der Weiterbildungsorganisationen mehrere Hinweise darüber, welche Herausforderungen die Organisationen in ihrer

Evaluationspraxis zu bewältigen haben, doch muss man berücksichtigen, dass die Selbstreporte von den Organisationen für den Zweck der erfolgreichen Qualitätstestierung verfasst worden sind. Für zahlreiche Organisationen haben die LQW-Testate eine existenzsichernde Bedeutung, denn die Einrichtungen müssen gegenüber Financiers einen Qualitätsnachweis erbringen. Organisationen werden daher verständlicherweise in den Selbstreporten ein positives Bild ihrer Einrichtung zeichnen wollen und es ist durchaus möglich, dass entscheidende Probleme der Evaluationspraxis verschwiegen werden. Zur Beantwortung der Fragestellungen 2 und 3 zu den Herausforderungen in der Evaluationspraxis und den zu ihrer Bewältigung benötigten Kompetenzen wurden deshalb insgesamt 24 Leitfadeninterviews mit Evaluationsverantwortlichen geführt, um diese explizit zu Herausforderungen zu befragen.

#### **4.3.1 Der Leitfaden und die Interviewdurchführung**

Da das wesentliche Ziel der Befragungen war, vergleichbare Aussagen der Interviewpartner/innen zu einem vorgegeben Themenkomplex zu erhalten, fiel die Wahl auf ein leitfadengestütztes Interview, dessen Vorteile für diesen Anwendungszweck auf der Hand liegen: „Der Vorteil eines Leitfadens gegenüber einem offenen narrativen Interview besteht also darin, dass die interessierenden Aspekte auch angesprochen werden und insofern eine Vergleichbarkeit mit anderen Interviews, denen der gleiche Leitfaden zugrunde lag, möglich ist“ (Marotzki, 2003, S. 114).

Die Datenerhebung wurde in Zusammenarbeit mit der damaligen Diplomandin Julia Busch realisiert, welche die Verwendung von Evaluationsergebnissen in Weiterbildungsorganisationen anhand von 15 der 24 Interviews untersucht hat (vgl. Busch, 2009). Zur Entwicklung des Leitfadens wurden nach Festlegung der Forschungsfragen konkrete Fragen zu drei Themengruppen gesammelt: (1) Herausforderungen und Entwicklungspotenziale, (2) Nutzung und Nutzen der Evaluation sowie (3) Kompetenzen und Know-how. Als Befragungsdauer wurden 30 Minuten angesetzt. Um den Leitfaden hinsichtlich Verständnis, Dramaturgie, Provozierung gehaltvoller Antworten und Zeitbedarf zu testen, wurde ein Probe-Interview mit einem Organisationsvertreter einer Volkshochschule durchgeführt, der mir persönlich bekannt war. Julia Busch und ich führten das Interview gemeinsam, um uns beide mit dem Leitfaden vertraut zu machen. Das Probeinterview ließ sich problemlos in das Sample mit aufnehmen, da es während der Interviewdurchführung zu keinen Problemen kam, keine Fragen mehr im Leitfaden ergänzt wurden, sondern der Leitfaden nur

geringfügig vereinfacht und die darin enthaltenen Nachfragehinweise präzisiert wurden.

Alle weiteren Interviews wurden telefonisch entweder von Julia Busch oder mir geführt und – wie auch das Probeinterview – digital aufgezeichnet. Die ersten drei Interviews hörten wir ausschnittsweise ab, um die Sinnhaftigkeit von Fragen zu kontrollieren und Verbesserungshinweise für unsere Interviewführung<sup>57</sup> abzuleiten. Es zeigte sich, dass die Frage danach, vor welchen Herausforderungen andere Weiterbildungsorganisationen bezüglich ihrer Evaluationspraxis stehen, zu schwer zu beantworten war. Die Frage wurde deshalb gestrichen, so dass der endgültige Leitfaden insgesamt 17 Fragen umfasste. Er wird im Folgenden Themenblock für Themenblock vorgestellt und erläutert. Nachfragen, die möglichst gestellt werden sollten, wenn die befragten Organisationsvertreter/innen zu den interessierenden Themen nicht von selbst Angaben machten, stehen in eckigen Klammern. Im Anhang B) ist der vollständige Leitfaden, einschließlich der einleitenden Worte, abgedruckt.

#### *Themenblock 1: Herausforderungen und Entwicklungspotenziale*

1. Wenn Sie mal an den Qualitätsbereich 5 von LQW zur Evaluation der Bildungsprozesse denken. Wo sehen Sie in Ihrer Organisation Herausforderungen, Unsicherheiten oder Schwierigkeiten in Bezug auf die Evaluation der Bildungsprozesse?  
[Gibt es noch etwas Weiteres, wo Sie Herausforderungen oder Schwierigkeiten sehen? (ggf. 2x nachfragen)]
2. Ich nenne Ihnen nun einige Bereiche, in denen Probleme bei der Evaluation auftreten können. Ich bitte Sie, jeweils kurz zu sagen, wie es in dem Bereich in Ihrer Organisation aussieht:
  - a) Akzeptanz der Evaluation durch Lehrende, Teilnehmende oder Mitarbeitende
  - b) Erhebungsinstrumente und -verfahren
  - c) Erhebung von Daten
  - d) Auswertung von Daten
  - e) Arbeitsaufwand
3. Welche von den Unsicherheiten oder Schwierigkeiten, die Sie genannt haben, stellt für Sie das größte Problem dar?  
[Und warum?]
4. Inwieweit ist in Ihrer Organisation beabsichtigt, die derzeitige Evaluationspraxis zukünftig zu verändern?  
[Sind Qualitätsentwicklungsziele oder strategische Entwicklungsziele im Rahmen von LQW definiert, in denen die Evaluation der Bildungsprozesse eine Rolle spielt?]
5. Inwieweit sehen Sie in Ihrer Organisation Unterstützungsbedarf bei der Evaluation von Bildungsprozessen, z.B. durch externe Personen?

---

57 Die Interviewführung orientierte sich an Hinweisen von Hermanns (2000).

Der erste Block konzentrierte sich auf die Herausforderungen, welche die Einrichtungen bezüglich ihrer Evaluation zu meistern haben. Während die erste Frage zunächst einen Bezug zum Qualitätsbereich 5 von LQW herstellt und die ersten drei Fragen direkt Herausforderungen thematisieren, fokussieren die Fragen 4 und 5 geplante Veränderungen und einen möglichen Unterstützungsbedarf, um indirekt auf Herausforderungen schließen zu können.

### *Themenblock 2: Nutzung und Nutzen der Evaluation*

6. Mit welchem Ziel werden in Ihrer Organisation eigentlich die Bildungsprozesse evaluiert?
7. Welche Ergebnisse erhält Ihre Organisation durch die Evaluation der Bildungsprozesse? In welcher Form liegen die Evaluationsergebnisse vor? [Wie kommen Sie von den Daten zu Konsequenzen? Welche Rolle spielen Bewertungskriterien und Sollwerte? Wie kommen Sie von den Daten zur Umsetzung?]
8. Wie werden in Ihrer Organisation die Evaluationsergebnisse genutzt?
9. Wer wird über die Ergebnisse der Evaluation informiert? [Teilnehmende, Dozenten, Mitarbeiter/innen, Leitung, Träger, Öffentlichkeit, Wissenschaft; Warum werden diese Personen informiert? Wie? (Internet, Zeitung, Bericht etc.?)]
10. Sie haben ja jetzt verschiedene Aspekte genannt, wie Sie die Evaluation und ihre Ergebnisse nutzen. Wenn Sie so insgesamt auf die Evaluation der Bildungsprozesse schauen: Wie beurteilen Sie den Nutzen der Evaluation insgesamt? [Wie beurteilen Sie den Aufwand für die Evaluation im Verhältnis zum Nutzen?]
11. Was meinen Sie: Wo müsste man ansetzen, wenn man den Nutzen der Evaluation für Ihre Organisation erhöhen wollte?

Der zweite Themenblock umfasste Fragen, die insbesondere Informationen über die Verwendung von Evaluationsergebnissen für die Diplomarbeit von Busch (2009) liefern sollten. Allerdings waren mehrere Fragen so angelegt, dass sie zum einen auch Informationen über Herausforderungen in der Evaluationspraxis hervorbringen, wie etwa die Frage nach der Nutzenerhöhung, und zum anderen interessante Zusatzinformationen liefern, wie etwa die Frage nach dem Ziel, mit dem Bildungsprozesse evaluiert werden.

### *Themenblock 3: Kompetenzen und Know-how*

- |   |
|---|
| 12. Welche Kompetenzen braucht man aus Ihrer Sicht, um Bildungsprozesse effizient zu evaluieren?  |
| 13. Welche dieser Kompetenzen sind in Ihrer Organisation vorhanden?<br>[Wie haben Sie diese Kompetenzen erlangt?]   |
| 14. Mal angenommen, Sie hätten Zeit und Geld, ein Seminar zum Thema „Evaluation von Bildungsprozessen“ zu besuchen. Was würden Sie oder eine Person aus Ihrer Organisation dort lernen wollen? Was wäre für Sie bzw. für Ihre Organisation hilfreich? |
| 15. Was verstehen Sie persönlich eigentlich unter Evaluation? Wie würden Sie also Evaluation definieren?  |
| 16. Wenn ich andere Personen in Ihrer Organisation fragen würde, was sie unter Evaluation verstehen: Würden diese Personen ähnlich antworten?   |

Der dritte Themenblock diente zunächst der Beantwortung der dritten Forschungsfrage zu den Kompetenzanforderungen und thematisierte die Selbsteinschätzung der Befragten, welche Kompetenzen für eine effiziente Evaluation von Bildungsprozessen nötig seien. Die hypothetische Frage zu den gewünschten Lerninhalten sollte einerseits Hinweise auf Kompetenzanforderungen liefern, andererseits aber auch indirekte Hinweise auf Herausforderungen hervorbringen, denn die interviewten Organisationsvertreter/innen werden vermutlich die Lerninhalte benennen, in denen Herausforderungen in der Organisation bestehen. Die Fragen 15 und 16, dienten dazu, das Evaluationsverständnis zu erheben, weil davon ausgegangen wurde, dass dieses – gewissermaßen als intervenierender Faktor – in einem Zusammenhang zur Evaluationspraxis steht.

### *Abschluss*

- |  |
|--|
| 17. Gibt es noch etwas Wichtiges aus Ihrer Sicht zu sagen? |
|--|

Die letzte Frage diente dazu, das Interview nicht abrupt zu einem Ende kommen zu lassen. Die Abschlussfrage sollte es den Interviewten ermöglichen, den Themenkomplex noch einmal in der Gänze zu betrachten und Aspekte zu nennen, nach denen nicht gefragt wurde.

Um gleiche Ausgangsbedingungen für alle Interviews zu schaffen, wurde der Interviewleitfaden den Organisationsvertreter/innen nicht vorab zugestellt, auch wenn einige ausdrücklich danach gefragt hatten. Die Interviewdurchführung verlief reibungslos. Lediglich bei einem Interview fiel es der Interviewperson schwer, die Fragen auf die eigene Praxis zu beziehen, was zu einer etwas zähen und angespannten Gesprächssituation führte, doch konnten wertvolle Informationen für die Beantwor-



tung der Forschungsfragen gewonnen werden. Interessant war, dass einzelne Befragte Schwierigkeiten mit dem Begriff „Problem“ hatten, von dem sie sich dezidiert abgrenzten: „Ähm, ja. Also nein ich sehe keine Probleme und Schwierigkeiten. Das ist auch für mich so ein, wenn ich das Wort Problem höre, da stehen mir manchmal, ja dann, es sind Aufgaben (lacht). Probleme, Probleme sind immer auf den Kopf gestellte Aufgaben“ (Interview 11, Absatz 53).

#### 4.3.2 Der begleitende Online-Fragebogen

Neben dem Interviewleitfaden wurde ein größtenteils standardisierter Online-Fragebogen entwickelt, den die Befragten vor dem Interview ausfüllen sollten.<sup>58</sup> Der Bogen diente dazu, zusätzliche Angaben über die Organisation der Befragten und deren Funktion innerhalb dieser Organisation zu erhalten und das Interview um diese Informationen zu entlasten. In drei Fällen haben die Befragten den Bogen erst nach dem Interview ausgefüllt, weil der erste Anruf, um einen Interviewtermin zu vereinbaren, gleich in ein Interview mündete oder weil die Interviewpersonen das Ausfüllen vergessen hatten.

Der Bogen enthielt zum einen Fragen nach Umfang, Instrumenten und den beteiligten Personen der in den Organisationen durchgeführten Evaluationen von Bildungsprozessen. Zum anderen wurden Organisationsmerkmale erfasst, bei denen ein Unterschied hinsichtlich der Evaluationspraxis zu vermuten war, etwa der Art des Bildungsangebots oder der Größe<sup>59</sup> der Organisation. Die Online-Umsetzung des Fragebogens erfolgte mit dem kostenfreien Tool „LimeSurvey“, dessen Einsatz Kuckartz et al. (2009) beschreiben. Die Person, die auch für den Pretest des Leitfadens interviewt wurde, füllte den Bogen mithilfe der Methode des lauten Denkens (vgl. Häder, 2010, S. 393-394) aus, um ihn auf Verständlichkeit und Beantwortbarkeit zu testen. Aufgrund der Ergebnisse des Pretests wurden zwei Fragen zur Anzahl von Unterrichtsstunden und Teilnehmenden gestrichen, da sie ad hoc (ohne vorliegende Unterlagen) zu schwer zu beantworten waren. Der endgültige Fragebogen ist in Abb. 33 wiedergegeben.

---

58 Standardisierte Fragebögen mit Interviews zu kombinieren ist ein übliches Verfahren, so sieht beispielsweise das problemzentrierte Interview nach Witzel (2000) den Einsatz eines Kurzfragebogens für sozialstatistische Daten vor.

59 Käßlinger (2009) berichtet beispielsweise über Unterschiede beim Bildungscontrolling in Abhängigkeit der Organisationsgröße und auch bei den Wirkungen von LQW findet sich diese Variable als relevantes Unterscheidungsmerkmal (vgl. Rädiker, 2006, 2010b).

Abb. 33: Der begleitende Online-Fragebogen

1. Bitte geben Sie den Namen Ihrer Organisation an: _____
2. Wie viele Veranstaltungen wurden in Ihrer Organisation im Jahr 2008 in etwa durchgeführt? (Bitte werten Sie Angebote mit mehreren Terminen als eine Veranstaltung). Veranstaltungen im Jahr 2008: _____
3. Wie viele Mitarbeiter/innen arbeiten derzeit in Ihrer Organisation? (Gemeint sind NICHT Personalstellen, sondern die Anzahl der Mitarbeiter/innen) hauptamtliches pädagogisches Personal ca.: _____ Verwaltungspersonal ca.: _____ nebenberuflich, freiberuflich und ehrenamtlich Lehrende ca.: _____
4. Bitte wählen Sie aus der Liste die beiden Hauptfinanzquellen Ihrer Einrichtung aus. o Teilnehmende/Selbstzahler                      o Kommunen, Länder, Bund, EU o Betriebe/Auftraggeber                          o Träger Ihrer Einrichtung o Arbeitsagentur (z.B. Bildungsgutscheine)      o Sonstige
5. Seit wie vielen Jahren wird in Ihrer Organisation das LQW-Modell eingesetzt? Seit _____ Jahr(en)
6. Hat Ihre Organisation eine AZWV-Anerkennung? <sup>60</sup> o ja o nein
7. Bitte beschreiben Sie kurz die inhaltlichen Schwerpunkte des Bildungsangebots Ihrer Organisation.
8. Liegt der Schwerpunkt der Bildungsangebote auf der beruflichen Bildung? o ja o nein
9. Bitte beschreiben Sie kurz die hauptsächlich angebotenen Veranstaltungsformen (Abendkurse, Bildungsur-laube, Wochenendseminare, ...).
10. Bitte beschreiben Sie kurz die Zielgruppen, die Ihre Organisation hauptsächlich anspricht.
11. Welche Aspekte des Bildungsprozesses werden in Ihrer Organisation evaluiert? (stichpunktartig)
12. Wie häufig und in welchem Umfang werden in Ihrer Organisation die Bildungsprozesse evaluiert? (stichpunktartig)
13. Welche Instrumente kommen bei der Evaluation der Bildungsprozesse in Ihrer Organisation zum Einsatz? (stichpunktartig)
14. Welche Personen (z.B. päd. MA, Verwaltungskräfte, Leitung, ...) sind an der Durchführung der Evaluation beteiligt und welche Aufgabe übernehmen diese Personen jeweils (z.B. Datenerhebung, Auswertung, Um-setzung, ...)?
15. Kreuzen Sie bitte an, welche Funktion Sie in Ihrer Organisation inne haben. (Mehrfachnennungen möglich) o Qualitätsbeauftragte/r                      o Leiter/in o päd. Mitarbeiter/in                          o Verwaltungsleiter/in o Verwaltungsmitarbeiter/in                  o Sonstiges: _____
16. Beinhaltet Ihre Stellenbeschreibung den Aufgabenbereich Evaluation? o ja o nein

60 AZWV steht für „Anerkennungs- und Zulassungsverordnung Weiterbildung“. Unter [www.azwv.de](http://www.azwv.de) finden sich weiterführende Informationen der Bundesagentur für Arbeit. Zum 01.04.2012 wurde AZWV umbenannt in AZAV „Akkreditierungs- und Zulassungsverordnung Arbeitsförderung“.

### 4.3.3 Das Interview-Sample

Es wurde eine quotierte Auswahl vorgenommen: Von jedem der ersten fünf<sup>61</sup> in Abschnitt 4.2.1 beschriebenen Organisationstypen wurden zunächst fünf Interviews eingeplant, weil bei einer Zufallsauswahl übermäßig viele Volkshochschulen im Sample dabei gewesen wären, bei denen man zumindest in Ansätzen davon ausgehen kann, dass sie aufgrund ihres vergleichsweise ähnlichen Bildungsangebots auch ähnlichen Problemen in der Evaluationspraxis begegnen. Da der ursprüngliche fünfte Organisationstyp der „wirtschaftsnahen Organisationen“ im Prozessverlauf mit den kommerziell orientierten zusammengelegt wurde, liegen für diese Gruppe neun Interviews vor. Die Abb. 34 informiert darüber, wie viele Organisationen angefragt wurden und wie viele teilgenommen haben. Es gab keine systematischen Ausfälle; am häufigsten meldeten Organisationen zurück, keine Zeit zu haben, oder reagierten nicht auf die Anfrage. Mehrere Organisationen beteiligten sich nicht, weil sie sich nicht weiter nach LQW testieren lassen wollten; die Anzahl dieser Organisationen ist in Klammern vermerkt.

**Abb. 34: Anzahl angefragter und teilgenommener Organisationen**

	sozial/ kulturell	weltanschaulich gebunden	Volkshoch- schulen	kommerziell/ wirtschaftsnah	Gesamt
angefragt	9	12 (+2)	10	12 (+3)	43 (+5)
teilgenommen	5	5	5	9	24

\* In Klammern ist die Anzahl an Organisationen angegeben, die angefragt wurden, sich aber nicht mehr nach LQW testieren lassen.

Die Kontaktierung der Organisationen erfolgte vorwiegend per E-Mail, nachdem die geeignete Ansprechperson per Anruf ermittelt wurde. Im Anhang C) ist das E-Mail-Anschreiben abgedruckt. Die interviewten Organisationsvertreter/innen hatten unterschiedliche Funktionen in ihren Organisationen inne, wobei sich bei der späteren Analyse keine nennenswerten Unterschiede bezüglich dieses Faktors zeigten. Die Funktionen können zusammen mit weiteren Merkmalen der Abb. 35 entnommen werden. Die Interviews wurden von Ende April bis Mitte Juli 2009 durchgeführt. Ihre durchschnittliche Länge betrug 32 Minuten, wobei wenige sehr viel kürzere, aber mehrere längere Interviews dabei waren. Das kürzeste Interview hatte eine Dauer von 16 Minuten, das längste dauerte 52 Minuten, wobei in dieser Zeit auch der Online-Fragebogen ausgefüllt wurde.

61 (1) sozial/kulturell organisiert, (2) weltanschaulich gebunden, (3) Volkshochschulen, (4) kommerziell orientiert, (5) wirtschaftsnah

Abb. 35: Übersicht über das Interview-Sample

	Organisations- größe***	Veranstaltungen pro Jahr	Schwerpunkt berufliche Bil- dung	AZVV- Anerkennung	LOW-Modell in Jahren	Funktion(en) der Interviewperson
<b>sozial/kulturell</b>						
I02*	mittel	400	nein	ja	7,0	Leitung
I10	klein	200	ja	nein	5,0	Leitung + QB**
I11	klein	60	ja	nein	1,0	Leitung + QB + päd. MA**
I21	klein	32	ja	nein	2,0	QB
I24	mittel	200	nein	ja	3,0	Sonstige
<b>weltanschaulich gebunden</b>						
I03	groß		ja		1,0	Leitung + QB + päd. MA
I09	klein	548	nein	nein	4,0	QB + päd. MA
I15	mittel	1971	ja	nein	5,0	QB + päd. MA
I19	klein		ja	ja	3,0	QB + päd. MA
I23	mittel	60	nein	nein	5,0	Leitung
<b>Volkshochschulen</b>						
I01	klein	1.075	nein	nein	5,0	päd. MA
I06	mittel	1.536	nein	nein	1,0	päd. MA
I12	groß	4.700	nein	nein	8,0	QB + päd. MA
I17	klein	1.200	nein	ja	5,0	Leitung + QB + päd. MA
I22	mittel	603	nein	ja	4,0	päd. MA
<b>kommerziell/ wirtschaftsnah</b>						
I04	groß	96	ja	ja	5,0	QB + Verwaltung
I05	klein	80	ja	ja	1,0	Leitung + päd. MA
I07	mittel	25	ja	ja	4,0	QB
I08	groß	30	ja	ja	4,0	Leitung
I13	groß		ja		4,0	Leitung
I14	klein	200	ja	nein	1,5	Leitung
I16	klein	180	nein	nein	3,0	Leitung
I18	mittel	300	nein	ja	3,0	Leitung + QB
I20	groß	54	ja	ja	3,0	Leitung

\* Die Interviews wurden entsprechend der Reihenfolge ihrer Durchführung nummeriert.

\*\* QB = Qualitätsbeauftragte/r, päd. MA = pädagogische/r Mitarbeiter/in; Geschäftsführung und stellvertretende Leitung wurden als „Leitung“ eingestuft.

\*\*\* Organisationsgröße: klein = unter 5; mittel = 5-15; groß = über 15 hauptamtliche päd. Mitarbeitende

#### 4.3.4 Transkription der Interviews

Alle Interviewaufzeichnungen wurden verschriftlicht. Da in die spätere Auswertung im Wesentlichen nur der Inhalt der Interviews (und keine verbalen Besonderheiten etc.) einfließen sollten, wurde ein einfaches Transkriptionssystem in Anlehnung an Kuckartz et al. (2008, S. 27-28) herangezogen:

1. Es wird wörtlich transkribiert, also nicht lautsprachlich oder zusammenfassend. Vorhandene Dialekte werden nicht mittranskribiert.
2. Die Sprache und Interpunktion wird leicht geglättet, das heißt an das Schriftdeutsch angenähert.
3. Deutliche, längere Pausen werden durch Auslassungszeichen in Klammern (...) markiert.
4. Zustimmungende bzw. bestätigende Lautäußerungen der interviewenden Person („Mhm“, „Aha“ etc.) werden nicht mittranskribiert, sofern sie den Redefluss der befragten Person nicht unterbrechen.
5. Einwürfe von anderen Personen in den Redefluss („Okay“, „Ja“ etc.) werden in Klammern gesetzt.
6. Auffällige Lautäußerungen wie lachen oder seufzen werden in Klammern notiert.
7. Die Aussagen der interviewenden Person werden durch „I:“, die der befragten Person durch ein „B:“ eingeleitet.
8. Nach einem Sprecherwechsel wird eine Leerzeile eingefügt, um die Lesbarkeit zu erhöhen.
9. Ereignisse während des Interviews werden in eckigen Klammern ergänzt, z.B. [spricht mit Kollegin].
10. Nicht verständliche Abschnitte werden mit zwei Fragezeichen in eckigen Klammern [??] gekennzeichnet.

Die Transkription wurde von einer externen, geschulten und erfahrenen Kraft übernommen. Die Passung von Interview und Transkript wurde von mir stichprobenartig überprüft. Bei zwei Interviewaufnahmen war die interviewende Person aus technischen Gründen schwer verständlich, jedoch konnten die gestellten Fragen mithilfe des Leitfadens rekonstruiert werden. Fehler in der Orthographie wurden größtenteils im Transkript verbessert. In der vorliegenden Arbeit abgedruckte

Textpassagen wurden noch einmal hinsichtlich Rechtschreibung und Kommasetzung kontrolliert und korrigiert. Im Anschluss an die Transkription wurden alle Angaben, die einen Rückschluss auf eine Organisation oder Person erlauben, anonymisiert. Dabei wurden die betreffenden Informationen, wenn möglich, durch eine Formulierung auf abstrakter Ebene ersetzt und in eckige Klammern gesetzt, z.B. [Bundesverband], [Abteilungsleiterin], [Organisationsname], [Städtename] etc.

#### **4.3.5 Die qualitative Inhaltsanalyse**

Wie die Selbstreporte wurden die Interviewtranskripte mithilfe der inhaltlich strukturierenden qualitativen Inhaltsanalyse ausgewertet, deren Ablauf detailliert im Abschnitt 4.2.2 beschrieben wurde. Die Definition der Hauptkategorien erfolgte inhaltlich und chronologisch entlang des Interviewleitfadens, wobei mit Ausnahme der letzten Frage eine, zum Teil auch zwei, der Fragen in eine Kategorie übersetzt wurden. Zur Erhöhung der Codierqualität wurden die ersten drei Texte von der Diplomandin Julia Busch und mir unabhängig voneinander mit den Hauptkategorien codiert und hinsichtlich Übereinstimmungen geprüft. Wie bei der Codierung der Selbstreporte zeigte sich hierbei eine sehr hohe Übereinstimmung, die Nicht-Übereinstimmungen waren eher technischer Art und betrafen beispielsweise den Umfang einer Codierung, nicht aber deren Inhalt. Die Codierungen erfolgten ebenfalls nach den oben aufgestellten Regeln: Doppelcodierungen werden innerhalb eines Interviewtranskripts vermieden und inhaltliche Doppelungen durch Memos gekennzeichnet; die codierten Textstellen müssen auch außerhalb ihres Kontextes verständlich sein.

Die weitere Codierarbeit erfolgte in zwei Phasen. In der ersten Phase ordnete die Diplomandin in allen weiteren Texten die Hauptkategorien zu und erarbeitete für das Datenmaterial ihrer Diplomarbeit auch Subkategorien zu ausgewählten Hauptkategorien. In der zweiten Phase wurden von mir alle Codierungen für die weitere Arbeit überprüft und angepasst. Für die Bildung von Subkategorien in den Hauptkategorien, die noch nicht von der Diplomandin bearbeitet worden waren, wurde ebenfalls ein zweistufiges Verfahren eingesetzt, indem zuerst eine weitere studentische Hilfskraft eine Vorabsortierung vornahm, die von mir überprüft, angepasst und erweitert wurde. Als weitere qualitätssteigernde Maßnahmen bei der Kategorienbildung kamen im Wesentlichen die weiter oben genannten Verfahren zum Einsatz, z.B. wurden ebenfalls kurze Zusammenfassungen der codierten Textsegmente erstellt.

Alle Hauptkategorien wurden im Prozessverlauf mit Kategoriendefinitionen versehen und auch zahlreiche Subkategorien erhielten eine detaillierte Definition, insbesondere bei Abgrenzungsproblemen. In Abb. 36 sind die Definitionen für die Subkategorien der Hauptkategorie „Problemfelder der Evaluation“ aufgeführt; im Anhang D) finden sich zusätzlich die Definitionen für alle Hauptkategorien.

**Abb. 36: Definitionen der Subkategorien zur Hauptkategorie „Problemfelder der Evaluation“**

01.1 Instrumente, Verfahren	Probleme, die sich auf die eingesetzten Evaluationsinstrumente und Evaluationsverfahren (=LQW-Terminologie) beziehen. Abgrenzung: Die Datenerhebung gehört nicht zum Evaluationsverfahren.
01.2 Datenerhebung	Probleme, welche die Datenerhebung im engeren Sinn betreffen: Verteilung und Ausfüllen von Fragebögen, Rückläufe etc.
01.3 Akzeptanz, Belastung	Aussagen, die a) Akzeptanzprobleme erkennen lassen, z.B. Widerstände, Vorbehalte oder b) auf Belastung hinweisen, z.B. zu viele, zu häufige, zu schwere Erhebungen. Der Code wird nur vergeben, wenn die Befragten es auch als Problem sehen. Der Code wird nicht vergeben, wenn gesagt wird, dass es Personen oder Personengruppen gebe, bei denen keine Akzeptanz da ist, das aber insgesamt gesehen kein Problem darstelle.
01.4 Datenauswertung und -bewertung	Probleme, welche die Auswertung (Softwareeinsatz, Aggregation etc.) oder die Bewertung von Daten betreffen.
01.5 Aufwand, Ressourcen	Probleme bezüglich Arbeitsbelastung, Effizienz, Effektivität, Aufwand-Nutzen-Relation, Aufwand allgemein, Belastung
01.6 Nutzen, Erkenntnisgewinn, Aussagekraft	Probleme, die den Nutzen der Evaluationspraxis betreffen, insbesondere Hinweise auf a) zu wenige, zu ungenaue oder zu wenig hilfreiche Informationen durch die Evaluation, b) aussageleere Daten, aber auch Probleme, die die Dimension c) Bewertung des Nutzens ansprechen.
01.7 Nutzung, Umsetzung, Ergebnisverwendung	Probleme zur Nutzung im Sinne von Umsetzung und Ergebnisverwendung.
01.8 Sonstiges	Probleme, die sich keiner vorherigen Kategorie zuordnen lassen und nicht häufiger als 4 Mal vorkommen. Nennungen, die mehr als einmal vorkommen, bekommen eine eigene Unterkategorie.
01.9 Bereiche ohne Probleme (mehr)	In diese Kategorie fallen sowohl Aussagen darüber, dass a) Probleme in bestimmten Bereichen nicht vorhanden sind, als auch, dass b) Probleme bewältigt wurden und nicht mehr existieren.

#### 4.4 Hinweise zum Umgang mit Quantifizierungen

Mehrere Publikationen widmen sich dem Thema Quantifizierungen in der qualitativen Sozialforschung, so etwa Grunenberg und Kuckartz (2010), die zahlreiche Einsatzmöglichkeiten deskriptiver Statistik vorstellen und unter anderem auch auf den Leitsatz von Seale (1999, S. 119) „Zähle das Zählbare“ verweisen. Quantifizierungen, meist in Form von Kategorienhäufigkeiten, spielten auch für diese Arbeit sowohl bei

der Analyse als auch bei der Ergebnisdarstellung eine bedeutende Rolle – eine Tatsache, die trotz der fortschreitenden Beliebtheit von Mixed-Methods-Ansätzen in der qualitativen Sozialforschung Skepsis und Irritationen hervorrufen kann und deshalb an dieser Stelle der Erläuterung Bedarf.

In dieser Arbeit wird Quantifizierungen ein anderer Stellenwert beigemessen, als es für die quantitative Forschung bei der Berechnung und Präsentation von Mittelwerten und anderen Kennzahlen etc. üblich ist. Während dort schnell die Zahlen dominieren und es auf Nachkommastellen ankommt, steht für diese Arbeit insbesondere die Funktion von Zahlen als Ordnungsrelation im Vordergrund: „Was kommt häufig vor, was kommt selten vor?“ sind hier die entscheidenden Fragen, denn es ist ein wichtiger Unterschied, ob bestimmte Inhalte nur einmal vorkommen und als Einzelmeinungen klassifiziert werden können oder ob ein Problem von nahezu allen Interviewpersonen in unterschiedlichen Facetten genannt wird. Der Vergleich von Kategorienhäufigkeiten für bestimmte Organisationstypen und andere Subgruppen hat einen heuristischen Wert und soll helfen, Muster und Regelmäßigkeiten zu identifizieren und zu illustrieren. Dabei werden die absoluten und relativen Häufigkeiten, insbesondere in Subgruppen, mit Bedacht gedeutet, denn eine Nennung kann bereits einen Anteil von 5 bis 10 Prozentpunkten ausmachen. Um Verzerrungen zu vermeiden, wurde sowohl in jedem Selbstreport als auch jedem Interview jeweils nur ausgewertet („gezählt“), ob ein Thema enthalten ist und nicht wie häufig. Aufgrund der größeren Stichprobe und der Zufallsauswahl ist bei der Analyse des Status quo deutlich weniger Zurückhaltung in puncto Quantifizierungen und Verallgemeinerungen geboten als bei der Analyse der Herausforderungen.

#### **4.5 Hinweise zur Verallgemeinerbarkeit der Ergebnisse**

Bei vielen Studien, insbesondere qualitativen, stellt sich die Frage, inwieweit die gefundenen Ergebnisse verallgemeinerbar sind.<sup>62</sup> Mayring (2007, Abs. 13) plädiert dafür offenzulegen, auf welche Form wissenschaftlicher Aussage generalisiert wird und auf welche Weise dies geschieht. Dies soll an dieser Stelle für die vorliegende Arbeit geschehen, wobei zwischen den beiden Datenquellen, Selbstreporte und Leitfadeninterviews, unterschieden werden muss.

---

<sup>62</sup> Es gibt auch generelle Kritik gegenüber der Generalisierung in der qualitativen Forschung; zur kurzen Übersicht siehe Mayring (2007, Abs. 9-12).



Für die Ermittlung des Status quo der Evaluationspraxis (Ergebnisse in Kapitel 5) wurde eine Zufallsstichprobe von 50 Selbstreporten gezogen, und zwar mit dem ausdrücklichen Ziel, die Ergebnisse auf die zugrundeliegende Grundgesamtheit der LQW-Organisationen in Deutschland zu verallgemeinern – zumindest auf die vier unterschiedlichen Organisationstypen, die in der Stichprobe vertreten waren. Aufgrund des gewählten Designs können die typischen und häufig zum Einsatz kommenden Instrumente und Verfahren der Evaluation genauso aufgezeigt werden wie die konkrete Nutzung der Evaluationsergebnisse. Auch wenn die gefundenen Unterschiede zwischen einzelnen Organisationstypen, Organisationsgrößen etc. bezüglich der Evaluationspraxis mit Vorsicht interpretiert werden, lassen sich doch die gefundenen großen Linien generalisieren. Für die Ermittlung der Herausforderungen in der Evaluationspraxis (Ergebnisse in Kapitel 6) stellt sich die Lage etwas anders dar. Hier lag der Fokus stärker darauf, die Bandbreite zu explorieren, welche Probleme in der Evaluationspraxis auftreten, wobei die Verallgemeinerung kein primäres Ziel war. Dementsprechend wurden die Fälle nach Quoten unterschiedlicher Organisationstypen und nicht per Zufall ausgewählt. Trotz allem ist davon auszugehen, dass sich zumindest die *häufig* auftretenden Probleme größtenteils auf die Grundgesamtheit der LQW-Organisationen in Deutschland übertragen lassen. Denn zum einen stellte sich der Organisationstyp, der als Kriterium für die Stichprobensammlung herangezogen wurde, als nur geringer Einflussfaktor auf diese großen Problemlinien heraus. Und zum anderen ist es eher unwahrscheinlich, dass die anderen Weiterbildungsorganisationen, die häufig ähnlich strukturiert sind, ähnliche Bildungsangebote vorhalten und die LQW-Anforderungen im Bereich der Evaluation abgearbeitet haben, vor gänzlich anderen Problemen bei ihrer Evaluationspraxis stehen.

Als eine Strategie zur Gewährleistung und Überprüfung der Generalisierbarkeit nennt Mayring (2007, Abs. 16) die Triangulation im Sinne der Kombination verschiedener Studien (zum Triangulationsbegriff vgl. Flick, 2011). Auch in der vorliegenden Studie wurden die beiden Datenquellen nicht als hermetisch abgeschlossene und voneinander getrennte Einheiten betrachtet. Im Gegenteil wurden bei der Auswertung mehrere Querverbindungen und Bezüge zwischen den Selbstreporten und den Leitfadeninterviews hergestellt. Insbesondere bei den berichteten Problemen in der Evaluationspraxis fanden sich zahlreiche Übereinstimmungen, aber auch einige ergänzende bzw. relativierende Ergebnisse (vgl. Abschnitt 6.1).



## 5 Der Status quo der Evaluationspraxis

Das fünfte Kapitel liefert eine Übersicht über die Evaluationspraxis in Weiterbildungsorganisationen, die das LQW-Modell anwenden. Dafür werden die Ergebnisse einer Analyse von 50 Selbstbeschreibungen von LQW-Organisationen vorgestellt. Das bei der Analyse gewählte methodische Vorgehen, die inhaltlich strukturierende Inhaltsanalyse, wurde detailliert in Abschnitt 4.2.2 beschrieben. In diesem Kapitel stehen nicht die einzelnen Organisationen im Vordergrund, sondern die Gesamtheit aller untersuchten 50 Organisationen hinsichtlich ausgewählter thematischer Schwerpunkte. Die Darstellung dieser Schwerpunkte folgt dem prototypischen Ablauf einer Evaluation, wie er in Abschnitt 2.6 vorgestellt wurde. Dementsprechend widmet sich Abschnitt 5.1 dem Status quo bezüglich der Evaluationszwecke und den Evaluationsgegenständen, dem sich Abschnitt 5.2 zu Evaluationsverfahren, Evaluationsdesigns und zur Datenerhebung anschließt. In den Abschnitten 5.3 und 5.4 werden die Ergebnisse zu den Themen Datenauswertung und Datenbewertung sowie der Umgang mit den Ergebnissen berichtet. Das fünfte Kapitel endet mit einer Zusammenfassung und einem Fazit zum Status quo und stellt abschließend eine Sammlung an Fragen zur Reflexion der Evaluationspraxis in Weiterbildungsorganisationen vor, welche auf Basis der Analyse zusammengestellt wurden.

### 5.1 Evaluationszwecke und Evaluationsgegenstände

In 46 der 50 Selbstreporte<sup>63</sup> von Weiterbildungsorganisationen konnten Angaben zu den Evaluationszwecken identifiziert werden. Nahezu alle dieser Organisationen geben einen oder mehrere der folgenden vier Gründe an, warum und wozu sie Evaluationen durchführen:

1. Erfolgskontrolle und Zielüberprüfung (von 27 Organisationen genannt)
2. Verbesserung, Optimierung und Entwicklung (24)
3. Wissens- und Erkenntnisgenerierung (17)
4. (Lehr-)Lernunterstützung (12)

*Erfolgskontrolle und Zielüberprüfung:* Die Hälfte der untersuchten Organisationen möchte mithilfe der Evaluation überprüfen, ob und inwieweit sie definierte Ziele er-

---

<sup>63</sup> Korrekterweise müsste es heißen: In den „Abschnitten der Selbstreporte über den Qualitätsbereich 5 zur Evaluation der Bildungsprozesse“, da nur diese in die Auswertung eingeflossen sind. Um dieses Sprachungetüm zu vermeiden, wird meist nur von „Selbstberichten“ gesprochen.

reicht. Hauptsächlich geht es den Organisationen darum, zu kontrollieren, ob die eigenen Ansprüche und Versprechen gegenüber den Teilnehmenden erfüllt werden konnten. Die Ansprüche leiten sich größtenteils aus dem Leitbild oder der Definition gelungenen Lernens ab, wobei nur wenige Organisationen die Ansprüche konkret benennen. So ist es eher die Ausnahme, dass eine Organisation schreibt, sie wolle die im Leitbild formulierte „Teilnehmendenorientierung“ (445:2)<sup>64</sup> mithilfe der Evaluation beurteilen. Die Organisationen wollen überprüfen, ob bzw. in welchem Ausmaß gelungenes Lernen tatsächlich stattfindet und Ziele des Leitbildes umgesetzt werden konnten. Sie folgen damit direkt einem Aspekt aus der LQW-Definition zur Evaluation von Bildungsprozessen, in der „die Realisierung des eigenen institutionellen Anspruchs“ (Zech, 2006, S. 68) wörtlich enthalten ist. Neben der Überprüfung des eigenen Anspruchs wollen einige Organisationen dezidiert die Erreichung von Lernerfolgen sowie Lern- und Kurszielen beleuchten und gegebenenfalls eruieren, aus welchen Gründen Maßnahmeziele nicht erreicht werden konnten (z.B. 432:4). Evaluation soll aber auch dazu dienen, anhand von Daten über die Teilnehmendenstruktur zu kontrollieren, ob man die Personen erreicht, die man erreichen möchte (543:26) oder ob vorgenommene Änderungen an einer Maßnahme Wirkung zeigen (432:16). Zusammenfassend kann man sagen, dass es den Organisationen bezüglich Erfolgskontrolle und Zielüberprüfung um die „Vergewisserung“ (669:5) über die Wirksamkeit und den Erfolg des eigenen Handelns geht und Evaluation so zum „Gradmesser“ (317:16; 385:24) für Erfolg wird.

*Verbesserung, Optimierung und Entwicklung:* Die Organisationen wollen mithilfe der Evaluation verbessern, optimieren, entwickeln, anpassen und nachbessern – um einige typische Formulierungen aufzugreifen. Im Fokus der „kontinuierlichen und systematischen“ (669:7) Entwicklung und Verbesserung stehen bei den Organisationen z.B. die Lerninhalte, Lernmethoden, Unterrichtsmaterialien, Räumlichkeiten und die Organisation (317:16), das Programmangebot und die Kursverwaltung (303:6), die Inhalte und pädagogischen Konzepte (543:17) oder – abstrakt und allgemein, ohne weitere Hinweise – das Qualitätsniveau oder das Angebot (u.a. 252:3; 343:8). Man erhofft sich durch die Evaluation Anregungen und Inspiration für die Weiterentwicklung der Bildungsarbeit (721:9), wobei die meisten Organisationen auf die Mängel, Defizite und Probleme schauen, die durch die Evaluation zu Tage treten, und

---

64 Die Zahlen in Klammern geben in diesem Kapitel die Quelle der Zitate an: Die erste Zahl steht für die Nummer des Selbstreports, die zweite für die Absatznummer.

nur einzelne Organisationen betonen, dass sie positive Evaluationsergebnisse bei der Optimierung und Entwicklung berücksichtigen (u.a. 491:12).

*Wissens- und Erkenntnisgenerierung:* Laut mehrerer Aussagen dient die Evaluation dem Informationsgewinn, man möchte Daten und Aussagen vorliegen haben, einen systematischen oder vollständigen Überblick oder ein umfassendes Bild gewinnen, beispielsweise über die Eindrücke der Teilnehmenden oder die Beratungsanlässe. Besonders häufig steht dabei die Beobachtung der Teilnehmendenzufriedenheit vornan, nur zwei Organisation haben die Lehrenden bezüglich der Wissensgenerierung im Blick und möchten unter anderem eruieren, wie gut diese in die Organisation eingebunden sind und welche Ideen sie haben (479:93; 495:31). Die Wissens- und Erkenntnisfunktion von Evaluation steht zumeist isoliert da, denn nur wenige Einrichtungen fügen ihr einen weiteren Schritt an und formulieren zum Beispiel, dass die gewonnenen Informationen als Grundlage von Verbesserungen oder der Erfolgskontrolle dienen (339:45; 492:25).

*(Lehr-)Lernunterstützung:* Aussagen darüber, dass die Evaluation ausdrücklich die Lernenden unterstützen soll, sind in den Selbstreporten von etwa jeder vierten Organisation zu lesen. So soll die Evaluation vor allem unmittelbar die persönliche Entwicklung der Teilnehmenden unterstützen, individuelle Lernprozesse anregen, Rückmeldung über den Wissenstand liefern und die eigene Motivation steigern (211:35; 259:27; 312:35; 432:12). Evaluation ist hier also als „pädagogisches Instrument“ (492:24) zu werten. Doch auch mittelbar soll die Evaluation den Lernenden zu Gute kommen und zum Beispiel angemessene Lernsituationen nach Umsetzung der Evaluationsergebnisse erzeugen. Neben die Lernunterstützung tritt zudem die Lehrunterstützung: Die Lehrenden sollen durch die Evaluation einen Überblick über das Lerngeschehen sowie Anhaltspunkte für die Steuerung des Lehr-Lern-Prozesses erhalten (434:21; 436:23).

Über die vorgestellten vier Hauptzwecke hinaus werden zahlreiche weitere Evaluationszwecke beschrieben, allerdings von deutlich weniger Organisation. So heben einige hervor, dass Evaluation ihnen zur Reflexion, zur Bedarfsermittlung und Programmplanung, zur Legitimation und Außendarstellung sowie zum Informationsaustausch bzw. zur Kommunikation dient (u.a. 436:24; 491:12). Auch die Evaluation zu Steuerungszwecken, wie im Rahmen der allgemeinen Evaluation<sup>65</sup> und der Bil-

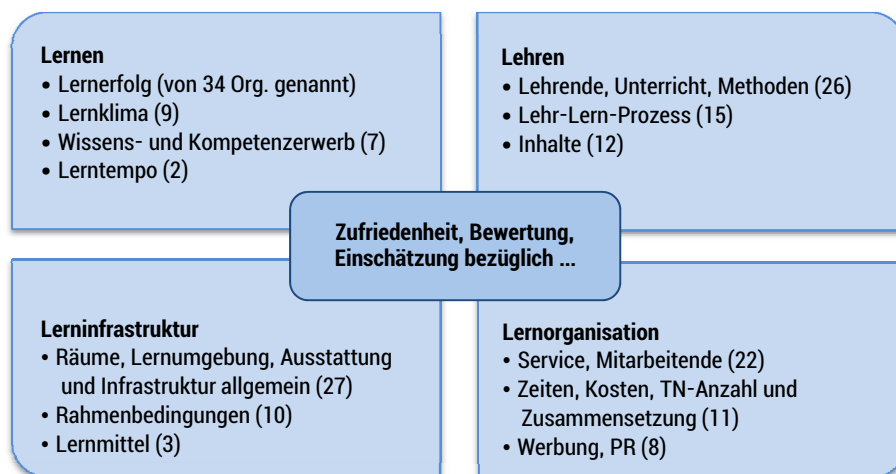
---

65 So befasste sich die Jahrestagung der Gesellschaft für Evaluation im Jahr 2009 ausführlich mit diesem Thema (vgl. [www.degeval.de](http://www.degeval.de)).

dungsevaluation derzeit diskutiert (vgl. Kuper, 2005, 2008), wird von einigen Organisationen auf mikrodidaktischer Ebene bezüglich der Lehr-Lern-Prozesse oder auf makrodidaktischer Ebene bezüglich der Gesamtorganisation benannt (u.a. 357:15; 731:47).

Die Selbstreporte wurden systematisch nach Aussagen durchgesehen, welche die Frage „Was evaluieren die Organisationen?“ beantworten, um die evaluierten Gegenstände und Inhalte sowie die leitenden Fragestellungen der Evaluationen zu eruieren. Man könnte vermuten, dass es sich bei den Gegenständen um Kurse, Maßnahmen, Seminare und andere Bildungsprogramme handelt, die in den Selbstreporten als evaluierte Gegenstände benannt werden. Dies ist jedoch nicht der Fall. Die Organisationen fassen Evaluationsgegenstände sehr viel kleinteiliger auf und die Grenze zu dem, was in der Evaluationsliteratur üblicherweise als Bewertungskriterien bezeichnet wird, ist teilweise fließend (vgl. hierzu ausführlich Abschnitt 2.4.2). Die häufigsten Nennungen zu Gegenständen, Inhalten und Fragestellungen lassen sich auf den in Kapitel 2 vorgestellten vier Ebenen des Bildungsprozesses (Lernen, Lehren, Lerninfrastruktur und Lernorganisation) verorten. Abb. 37 zeigt, welche Aspekte auf den einzelnen Ebenen identifiziert wurden, wobei weiter oben stehende Aspekte häufiger vorkommen.

**Abb. 37: Identifizierte Evaluationsgegenstände und Inhalte auf den vier Dimensionen Lernen, Lehren, Lerninfrastruktur und Lernorganisation**



Auf allen vier Ebenen kristallisiert sich jeweils ein Aspekt aufgrund der häufigeren Nennung deutlich heraus. So dominiert beim Lernen die Abfrage des Lernerfolges, wobei es jedoch häufig vage bleibt, was die Einrichtungen unter Lernerfolg konkret verstehen. Wenn der Lernerfolg von den Organisationen konkretisiert wird, dann einerseits im Sinne eines „weichen“ Lernerfolgs, der weit gefasst wird oder aus

Selbsteinschätzungen besteht („Lernerfolg bei der Veranstaltung“, 728:6, „Lernerfolg aus Sicht des Kunden“, 434:6), und andererseits als Lernerfolg im Sinne von „harten“ Leistungen, die durch Tests und Prüfungen nachgewiesen werden. Neben dem Lernerfolg fokussieren einige Organisationen auch ergänzend oder alternativ den Wissens- und Kompetenzerwerb; dabei handelt es sich vor allem um kommerziell orientierte bzw. wirtschaftsnahe Organisationen, seltener um Volkshochschulen.

Beim Lehren steht die Evaluation der Lehrenden bzw. ihrer Vorgehensweisen im Vordergrund. Ganz allgemein schreiben die Organisationen in diesem Zusammenhang von „Referenten“ und „Kursleitern“ sowie von „Unterricht“ und „Kursdurchführung“ (u.a. 211:16; 290:14; 339:24; 479:46). Wenn spezielle Aussagen zu den Lehrenden zu finden sind, dann geht es um ihre Fachkompetenz und Qualifikation, ihre pädagogische Arbeit und ihre Unterrichtsvorbereitung sowie ihre methodisch-didaktische Vorgehensweise. Erstaunlicherweise sind es nur 13 Organisationen, die explizit schreiben, dass die Evaluation auch die Kursinhalte, also den „Lehrstoff“ (325:9) berücksichtigt, z.B. ob der Kurs einen guten Einblick in das Themengebiet ermöglichte oder ob der Inhalt bewältigt werden konnte (731:19; 736:14). Vielleicht existiert bei den anderen Organisationen die Vorstellung, dass die Lernerfolge oder auch die Zufriedenheit ausreichender Beleg für gute bzw. angemessene Inhalte sind. Es sind übrigens vorwiegend die sozial/kulturell sowie die kommerziell orientierten bzw. wirtschaftsnahen Organisationen, welche die Inhalte benennen.

Bei der Lerninfrastruktur dominieren Aussagen zu den Lernbedingungen, insbesondere bezüglich der Räumlichkeiten und ihrer Sauberkeit sowie der Ausstattung. Hinzu kommen die Rahmenbedingungen, die jedoch selten konkretisiert werden, sowie bei sehr wenigen Organisationen auch die Lernmittel, wo es beispielsweise darum geht, die Präsentation und den Umfang des Lernmaterials abzufragen (434:14).<sup>66</sup>

Bei der Lernorganisation stehen an erster Stelle allgemeine Aussagen zum Service und zu den Mitarbeitenden. Konkretere Nennungen reichen von der Anmeldung über die Beratung bis hin zur Betreuung. Im Fokus der Evaluation stehen dabei nicht nur Teilnehmende, sondern auch Lehrende, die z.B. über die Zusammenarbeit mit den Mitarbeitenden einer Organisation und ihrer Freundlichkeit befragt werden.

---

66 Die Lernmaterialien hätten auch der Dimension des „Lehrens“ zugeordnet werden können, da sie häufig in den Verantwortungsbereich der Lehrenden fallen. Im LQW-Modell werden jedoch die Lernmaterialien in den Spezifikationen des Qualitätsbereichs „Infrastruktur“ aufgeführt, weshalb der Zuordnung zur Ebene „Lerninfrastruktur“ der Vorrang gegeben wurde.

Etwa jede fünfte Organisation evaluiert Aspekte der vorwiegend formalen Rahmenbedingungen eines Bildungsangebots, wie etwa die Lernzeiten, die Kursgebühren, die Gruppengröße und die Zusammensetzung der Teilnehmenden (u.a. 333:14; 357:6; 483:14). Zudem sind einige Organisationen daran interessiert, wie die Teilnehmenden vom Kursangebot erfahren haben und wie die Vorab-Informationen bewertet werden (u.a. 376:6; 244:41).

Teilweise quer, teilweise ergänzend zu den bisher genannten Bereichen liegt der Aspekt der Zufriedenheit, der bezüglich Evaluation im Bildungsbereich häufig diskutiert wird und auch als Maßstab in der Evaluationsdefinition des LQW-Leitfadens genannt wird (vgl. Abschnitt 3.2). Der Aspekt der Zufriedenheit nimmt eine gewisse Doppelstellung und Sonderrolle ein. Einerseits wird die „Zufriedenheit mit ...“ einem konkreten Gegenstand abgefragt, und zwar insbesondere bezüglich der Lerninfrastruktur und der Lernorganisation, weniger beim Lehren und kaum beim Lernen. Diese Überschneidung mit einem der anderen Bereiche soll durch die Überlappung des Bereichs „Zufriedenheit“ mit den anderen vier Bereichen in Abb. 37 verdeutlicht werden. Andererseits erhebt mehr als die Hälfte aller Organisationen zusätzlich und einige auch nur ausschließlich die Zufriedenheit der Teilnehmenden und vereinzelt auch der Lehrenden auf allgemeiner Ebene im Sinne einer „Overall-Satisfaction“ oder in Bezug auf ein ganzes Bildungsangebot. Bezüglich der Zufriedenheit ist zudem als besondere Auffälligkeit anzumerken, dass sie von mehreren Organisationen als Indikator für den Lernerfolg beschrieben wird (u.a. 495:29).

Die Definition gelungenen Lernens und zum Teil auch das gesamte Leitbild spielen für ein Drittel der Organisationen eine wichtige Rolle für die Bestimmung der zu evaluierenden Gegenstände. So betont eine Organisation, dass die Befragungsgegenstände direkt die eigene Definition gelungenen Lernens widerspiegeln würden (276:6). Der Grad der Operationalisierung ist dabei unterschiedlich ausgeprägt. Während der Großteil der Organisationen aus der Idealvorstellung gelungenen Lernens konkrete zu evaluierende Aspekte wie etwa die Kenntniserweiterung ableitet oder übernimmt, verzichtet eine Einrichtung sogar völlig auf eine Operationalisierung des gelungenen Lernens und fragt ihre Teilnehmenden direkt, ob das Lernen aus ihrer Sicht gelungen sei (434:15).

Zu den weiteren, von jeweils etwa sechs Organisationen genannten evaluierten Gegenständen gehören: die Beratung, der Verbleib, Bedarfe und Wünsche, Weiterempfehlung und Kundenbindung, die Gründe für den Kursabbruch und die Erfüllung von



Erwartungen. Zudem hebt eine Organisation hervor, nur die Gegenstände und Kriterien in der Evaluation zu berücksichtigen, die sie auch verändern kann (429:5).

In Abb. 38 sind die wesentlichen Evaluationszwecke und Gegenstände noch einmal für verschiedene Organisationstypen getrennt aufgeführt, wobei größere Abweichungen fett hervorgehoben wurden.<sup>67</sup> Unter den sozial bzw. kulturell orientierten Organisationen geben sehr wenige Einrichtungen an, Evaluation zur Verbesserung und Entwicklung oder zur Wissensgenerierung zu betreiben, während bei den kommerziell orientierten bzw. wirtschaftsnahen Organisationen das Gegenteil der Fall ist: Sie geben alle vier Zwecke überdurchschnittlich häufig an, wobei die Verbesserung und Entwicklung sowie die Wissensgenerierung dominieren. Bezüglich der Evaluationsgegenstände fällt auf, dass erstaunlich wenige weltanschaulich gebundene Organisationen angeben, das Lernen und Lehren zu evaluieren. Übrigens lassen sich zur Organisationsgröße weder Zusammenhänge zum Evaluationszweck noch zu den evaluierten Gegenständen finden.

**Abb. 38: Unterschiede bezüglich Zwecken und Gegenständen der Evaluation bei verschiedenen Organisationstypen**

	sozial/ kulturell	weltanschau- lich gebunden	Volkshoch- schulen	kommerziell/ wirtschaftsnah	Gesamt
	n = 10	n = 9	n = 20	n = 11	n = 50
<b>EVALUATIONSZWECKE</b>					
Erfolgskontrolle und Zielüberprüfung	6 (60%)	5 (63%)	9 (45%)	7 (58%)	27 (54%)
Verbesserung, Entwicklung	<b>2 (20%)*</b>	3 (38%)	11 (55%)	<b>9 (75%)*</b>	25 (50%)
Wissens- und Erkenntnisgenerierung	<b>1 (10%)</b>	3 (38%)	6 (30%)	<b>7 (58%)*</b>	17 (34%)
(Lehr-)Lernunterstützung	2 (20%)	3 (38%)	4 (20%)	3 (25%)	12 (24%)
<b>EVALUATIONSGEGENSTÄNDE</b>					
Lernen	8 (80%)	<b>3 (38%)*</b>	16 (80%)	10 (83%)	37 (74%)
Lehren	9 (90%)	<b>2 (25%)*</b>	17 (85%)	8 (67%)	36 (72%)
Lerninfrastruktur	7 (70%)	4 (50%)	16 (80%)	7 (58%)	34 (68%)
Lernorganisation	3 (30%)	5 (63%)	13 (65%)	4 (33%)	25 (50%)

Angegeben ist die Anzahl an Organisationen, die über die jeweiligen Zwecke bzw. Gegenstände berichtet haben. Die Prozentangaben beziehen sich auf die Anzahl an Organisationen des jeweiligen Typs. Beispielsweise berichten 60% der sozial/kulturellen Organisationen von Erfolgskontrolle und Zielüberprüfung.

**Fett** hervorgehoben sind Häufigkeiten, deren standardisiertes Residuum vom Betrag größer gleich 1 ist. Abweichungen, die nach Fishers-Exakt-Test statistisch signifikant sind, sind mit \* (p<5%) und \*\* (p<1%) gekennzeichnet.

<sup>67</sup> Als Hinweis für Auffälligkeiten wurden in den Tabellen dieses Kapitels solche Häufigkeiten fett hervorgehoben, deren standardisiertes Residuum vom Betrag her größer gleich 1 ist. Häufigkeiten, die so stark von den erwarteten Häufigkeiten abweichen, dass sie sich zudem nach dem Fishers-Exakt-Test als statistisch signifikant erweisen, sind zusätzlich mit Sternen gekennzeichnet. Der Fishers-Exakt-Test wurde für die hinter jeder Zelle liegende Vierfeldertafel berechnet. Zum generellen Umgang mit Quantifizierungen bei den qualitativen Ursprungsdaten vgl. Abschnitt 4.4.

## 5.2 Evaluationsverfahren, Evaluationsdesigns und Datenerhebung

Die Organisationen berichten am häufigsten davon, dass sie als Evaluationsverfahren<sup>68</sup> *Fragebögen* einsetzen. Hierbei überwiegen deutlich die Fragebögen für Teilnehmende, die fast alle Organisationen verwenden. Dabei ist es durchaus üblich, für verschiedene Bildungsbereiche auch verschiedene Bögen vorzusehen, so dass einzelne Organisationen von bis zu fünf unterschiedlichen Teilnehmenden-Fragebögen berichten (u.a. 259). Zum Teil kommen auch Fragebögen zum Einsatz, die sowohl vereinheitlichte Themenblöcke für alle Kurse als auch fortbildungsspezifische Themenblöcke umfassen (u.a. 343:15). Neben den Teilnehmenden werden auch die Lehrenden von knapp der Hälfte der Organisationen mithilfe eines Fragebogens um Rückmeldungen gebeten, denn das Wirken der Lehrenden bestimmt maßgeblich die Qualität der Einrichtung und sie nähmen eine zentrale Rolle im Bildungsprozess ein, wie eine Organisation nachdrücklich betont (308:4).

Die Fragebögen sind größtenteils standardisiert, nur gelegentlich finden sich Hinweise auf offene Fragen ohne Antwortvorgaben, die unter anderem der Erhebung von Wünschen dienen und von denen man sich gegenüber den geschlossenen Fragen weiterführende Informationen erhofft. Eine Organisation, die Angebote für benachteiligte Arbeitslose offeriert, hebt als Vorteil der offenen Fragen hervor, dass diese eine differenzierte Darstellung individueller Lernbedürfnisse ermöglichen und somit ein umfangreiches und detailliertes Feedback lieferten (325:19).

Dass der Einsatz von Fragebögen nicht ganz unproblematisch, sogar unangebracht sein kann, darauf weisen mehrere Organisationen hin. So könne beispielsweise (Persönlichkeits-)Bildung auch unbequem sein oder bildungsferne Teilnehmende könnten mit den Selbstlernphasen am Anfang eines Kurses überfordert sein, was einer häufigen bzw. zu Beginn stattfindenden Zufriedenheitsabfrage entgegenstehen würde (343:22; 428:3). Eine Organisation verwehrt sich sogar gänzlich gegen Fragebögen, da diese eine nicht gegebene Repräsentativität suggerierten, die Anwendung geschlossener Fragen für die angesprochenen Zielgruppen ausscheide und für die Auswertung offener Fragen keine Zeit vorhanden sei (436:17).

An zweiter Stelle der eingesetzten Evaluationsverfahren werden *Gespräche und mündliche Befragungen* genannt, die in unterschiedlichsten Variationen durchge-

---

68 Wie in Abschnitt 2.5 bereits angemerkt, werden die Begriffe „Verfahren“, „Instrumente“ und „Methoden“ in dieser Arbeit meist synonym verwendet. In den Auswertungskapiteln wird dem Terminus „Verfahren“ größtenteils der Vorzug gegeben wird, da dieser der LQW-Terminologie entspricht.

führt werden und denen mehrere Organisationen zahlreiche Vorteile attestieren. So seien Gespräche beispielsweise zeitsparend, aktuell und schnell, was dem Bedürfnis nach sofortiger Reaktionsmöglichkeit entgegenkomme. Dementsprechend könnten Konzepte laufender Kurse direkt angepasst und die Kundenbindung erhöht werden (259:23). Einige Organisationen weisen auch darauf hin, dass sie mit persönlichen Gesprächen Zielgruppen besser erreichten, etwa weil diese gegenüber dieser klassischen Kommunikationsform aufgeschlossener seien als gegenüber Fragebögen (u.a. 358:17).

Es lassen sich im Wesentlichen drei verschiedene Gesprächsformen identifizieren: Erstens finden gezielt und geplant Gespräche mit Teilnehmenden, Lehrenden bzw. pädagogischen Mitarbeitenden sowie Auftraggebenden statt. Zu diesen Gesprächen zählen insbesondere:

- klassische Auswertungsgespräche zwischen verantwortlichen Mitarbeitenden und Lehrenden,
- Einzelgespräche mit Teilnehmenden, vor allem in langfristigen Kursen kommerziell orientierter bzw. wirtschaftsnaher Anbieter,
- Gruppenbesprechungen mit Lehrenden und pädagogischen Mitarbeitenden, wie etwa Dozentenkonferenzen, sowie
- telefonische Abbrecherbefragungen.

Zweitens führen zahlreiche Organisationen Gespräche bzw. gesprächsaktivierende Methoden an, die direkt im Seminarprozess mit der Lerngruppe eingesetzt werden. Es handelt sich gewissermaßen um „Mini-Evaluationen“<sup>69</sup> (408:21), allen voran die Feedbackrunde, die vorwiegend zum Abschluss der Veranstaltungen durchgeführt wird. Die Lehrenden setzen laut Lehrplan regelmäßig Feedbackrunden ein oder wählen aus einem vorgegebenen Fundus an Evaluationsmethoden, wie Meinungsrauschen oder Gruppendiskussion, aus (543:17).

Schließlich sind drittens informelle, spontane Gespräche zu nennen, die zumeist mit Teilnehmenden stattfinden, etwa in der Pause, am Telefon oder als Rückmeldungen am Rande von Lehrveranstaltungen, Vorträgen oder Ausstellungseröffnungen. Mehrere Organisationen betonen die Wichtigkeit und Bedeutung dieser nicht formalisierten Verfahren, die „auf informellen Kontakten“ (290:3) basieren. Die Verfahren

---

69 Die Volkshochschule, die diesen Begriff verwendet, weist zugleich auf die Problematik der Dokumentation dieser Evaluationsverfahren hin.

wären etabliert und hätten sich bewährt, sie gehörten einfach zur „täglichen Praxis“ (408:20) und Teilnehmende, Lehrende und pädagogische Mitarbeitende seien teilweise persönlich miteinander bekannt.

Neben diesen beiden Hauptverfahren, den Fragebögen und Gesprächen, berichten jeweils knapp 20 der 50 Weiterbildungsorganisationen, dass sie *Tests und Prüfungen* sowie *Berichte* einsetzen. Bei den Tests und Prüfungen handelt es sich hauptsächlich um externe Prüfungen, z.B. von der Industrie- und Handelskammer oder im Rahmen des Europäischen Computerführerscheins, seltener um intern erstellte und durchgeführte Wissenstests oder mündliche Leistungskontrollen wie etwa Referate. Berichte nehmen im Rahmen der Evaluation eine Doppelfunktion ein: Einerseits dienen sie als Evaluationsverfahren, mit dessen Hilfe Bildungsveranstaltungen reflektiert werden, andererseits nutzen die Organisationen die Berichte zur Dokumentation. Es sind hauptsächlich die Lehrenden, die zumeist am Ende einer Veranstaltung einen Seminar- oder Kursbericht vorlegen, teilweise hierzu vertraglich verpflichtet worden sind. Derartige Berichte informieren meist über den Verlauf der Maßnahmen sowie aufgetretene Besonderheiten und Zwischenfälle und enthalten zahlreiche statistische Angaben über die Zusammensetzung der Teilnehmenden, Fehlzeiten und Eingliederungsquoten (u.a. 238:6). Vor diesem Hintergrund verwundert es nicht, dass Berichte besonders häufig von kommerziell orientierten bzw. wirtschaftsnahen Organisationen verwendet werden.

Nur etwa ein Drittel der Organisationen gibt an, auf bereits vorhandene Materialien bei der Evaluation zurückzugreifen. Bei diesen *Materialanalysen* handelt es sich vorwiegend um die Auswertung von Anmeldezahlen, Wartelisten, Teilnehmenden- und Anwesenheitslisten sowie anderen Statistiken. Als Evaluationsverfahren genannt, wenn auch deutlich seltener, werden schließlich noch die Hospitation und Beobachtung durch hauptamtliche pädagogische Mitarbeitende oder Ausbilder, die Erhebung von Anregungen und Beschwerden durch entsprechend ausgewiesene Postkästen mit eigens dafür erstellten Formularen sowie bei arbeitsmarktnahen Bildungsangeboten der Einsatz von Beurteilungsbögen und die Durchführung von Verbleibsstudien.

In Abb. 39 sind die Evaluationsverfahren aufgeführt, über die am häufigsten berichtet wurde. Zugleich ist der Darstellung zu entnehmen, wie sich diese Verfahren auf die unterschiedlichen Organisationstypen verteilen, wobei Auffälligkeiten im Sinne größerer Abweichungen von den (bei Unabhängigkeit) erwarteten Häufigkeiten fett hervorgehoben sind. In der Abbildung ist zunächst die breite Palette an Verfahren zu

erkennen, die bei der Evaluation zum Einsatz kommen und jeweils von mindestens einem Drittel der Organisationen angewendet werden. Bezüglich der einzelnen Verfahren fällt zudem auf, dass Fragebögen für Lehrende besonders in Volkshochschulen beliebt sind, während Gespräche im Seminarprozess in überdurchschnittlich vielen sozial bzw. kulturell orientierten Organisationen zum Einsatz kommen. Dies ist vermutlich auf die Art der Bildungsangebote in sozial bzw. kulturell orientierten Organisationen zurückzuführen, die durch Blockseminare an mehreren Tagen hintereinander zu häufig persönlichkeitsbezogenen Themen geprägt sind. Beachtenswert hoch ist auch die Bedeutung von informellen Gesprächen in Volkshochschulen, von denen die Hälfte dieses Organisationstyps berichtet. Die auffallend hohe Bedeutung der Berichte in kommerziell orientierten bzw. wirtschaftsnahen Organisationen ist durch die typischen Bildungsangebote im arbeitsmarktnahen Bereich zu erklären, die häufig mehrere Wochen oder Monate andauern und mit einem Bericht abgeschlossen werden. Kommerziell orientierte bzw. wirtschaftsnahe Organisationen sind häufiger unter den großen Organisationen anzutreffen und dementsprechend zeigt eine Zusammenhangsanalyse mit der Organisationsgröße, dass große Organisationen auch häufiger Berichte als Evaluationsverfahren einsetzen (Fishers-Exakt-Test;  $p = .037$ ;  $\Phi = .31$ ). Ansonsten zeigen sich keine Zusammenhänge zur Organisationsgröße.

**Abb. 39: Unterschiede bezüglich der Evaluationsverfahren bei verschiedenen Organisationstypen**

	sozial/ kulturell	weltanschau- lich gebunden	Volkshoch- schulen	kommerziell/ wirtschaftsnah	Gesamt
<b>EVALUATIONSVERFAHREN</b>	n = 10	n = 9	n = 20	n = 11	n = 50
Fragebögen für Teilnehmende	9 (90%)	8 (100%)	20 (100%)	11 (92%)	48 (96%)
Fragebögen für Lehrende	4 (40%)	2 (25%)	<b>12 (60%)</b>	4 (33%)	22 (44%)
Gespräche mit TN, Lehrenden, MA	7 (70%)	5 (63%)	12 (60%)	8 (67%)	32 (64%)
Gespräche im Seminarprozess	<b>8 (80%)**</b>	4 (50%)	<b>3 (15%)**</b>	5 (42%)	20 (40%)
Gespräche, informell/ungeplant	2 (20%)	<b>1 (13%)</b>	<b>10 (50%)</b>	4 (33%)	17 (34%)
Tests, Prüfungen	5 (50%)	2 (25%)	7 (35%)	5 (42%)	19 (38%)
Berichte	3 (30%)	3 (38%)	5 (25%)	<b>7 (58%)*</b>	18 (36%)
Materialanalysen	4 (40%)	2 (25%)	7 (35%)	3 (25%)	16 (32%)

Angegeben ist die Anzahl an Organisationen, die über die jeweiligen Verfahren berichtet haben. Die Prozentangaben beziehen sich auf die Anzahl an Organisationen des jeweiligen Typs.

**Fett** hervorgehoben sind Häufigkeiten, deren standardisiertes Residuum vom Betrag größer gleich 1 ist. Abweichungen, die nach Fishers-Exakt-Test statistisch signifikant sind, sind mit \* ( $p < 5\%$ ) und \*\* ( $p < 1\%$ ) gekennzeichnet.

Der Großteil der Weiterbildungsorganisationen beschränkt den Umfang der Evaluationsbemühungen gezielt auf ausgewählte Veranstaltungen bzw. Teilnehmende. Etwa die Hälfte der Einrichtungen konzentriert sich auf bestimmte Programm- oder

Fachbereiche, Kursreihen oder Themengruppen, wie etwa alle Kurse des Bereichs „Beruf und Karriere“, alle Kurse im Bereich „Deutsch als Fremdsprache“ oder alle Bildungsurlaube und Auftragsmaßnahmen. Zum Teil existieren auch externe Vorgaben, die eine Evaluation verpflichtend vorschreiben, wie etwa bei akkreditierten Bildungsangeboten für Lehrer/innen oder bei arbeitsmarktnahen Kursen (u.a. 479:79). Bei der Auswahl der evaluierten Veranstaltungen setzen mehrere Organisationen zudem auf die folgenden Strategien:

- Es werden die Veranstaltungen mit neuen Lehrenden, zu neuen Themen, an neuen Orten oder mit vielen Neukunden evaluiert (u.a. 620:7).
- Es werden nur Veranstaltungen mit einer vorgegebenen Minstdauer evaluiert (u.a. 388:9).
- Es werden die Veranstaltungen, die zuvor schlecht bewertet oder bei denen Schwierigkeiten bekannt bzw. vermutet wurden, evaluiert (u.a. 133:15).
- Es werden gezielt Veranstaltungen aus der Evaluation ausgeschlossen, zum Beispiel die Kurse im Bereich „Deutsch als Fremdsprache“ (u.a. 488:12).

Die genannten Strategien werden häufig mit Vorgaben zur Stichprobengröße, z.B. 10% aller Kurse oder mindestens 10 Kurse pro Fachbereich, und mit Rotationsverfahren kombiniert, um unter anderem sicherzustellen, dass innerhalb eines festgelegten Zeitraums alle Kurse oder alle Fachbereiche einmal evaluiert werden (u.a. 731:7). Organisationen, die eine Semesterstruktur bei ihren Angeboten verfolgen, evaluieren zumeist im halbjährlichen oder jährlichen und selten in einem längeren Rhythmus.

Etwa jede fünfte Organisation gibt an, flächendeckende Erhebungen durchzuführen. Es handelt sich dabei tendenziell um kleinere Organisationen, unter ihnen keine Volkshochschule, dafür aber zahlreiche sozial/kulturell orientierte Einrichtungen (vgl. Abb. 40). Nur selten wird berichtet, dass eine Organisation gezielt eine konkrete Fragestellung verfolgt und nur die für die Beantwortung der Fragestellung notwendigen Kurse oder Teilnehmenden befragt, z.B. herausfinden möchte, welche Wünsche Berufstätige haben und dabei nur die Teilnehmenden des Bereichs „Beruf und Karriere“ einbezieht (479:81).

**Abb. 40: Unterschiede bezüglich des Umfangs bei der Evaluation bei verschiedenen Organisationstypen und Organisationsgrößen**

	sozial/ kulturell	weltanschau- lich gebunden	Volkshoch- schulen	kommerziell/ wirtschaftsnah	Gesamt
alle Veranstaltungen	<b>6 (60%)**</b>	2 (29%)	<b>0 (0%)**</b>	3 (30%)	11 (24%)
gezielt ausgewählt	<b>4 (40%)**</b>	5 (61%)	<b>19 (100%)**</b>	7 (70%)	35 (76%)
Summe	10	7	19	10	46

	unter 5 HPM	5-15 HPM	über 15 HPM	Summe
alle Veranstaltungen	<b>6 (35%)</b>	3 (14%)	2 (18%)	11 (22%)
gezielt ausgewählt	10 (59%)	17 (77%)	8 (73%)	35 (70%)
Summe	16	20	10	46

Angegeben ist die Anzahl an Organisationen. Die Prozentangaben beziehen sich auf die Anzahl an Organisationen des jeweiligen Typs bzw. der jeweiligen Größe. In vier Selbstreports konnten keine Angaben bestimmt werden, weshalb sich die Gesamtanzahl der Organisationen von 50 auf 46 reduziert.

**Fett** hervorgehoben sind Häufigkeiten, deren standardisiertes Residuum vom Betrag größer gleich 1 ist. Abweichungen, die nach Fishers-Exakt-Test statistisch signifikant sind, sind mit \* ( $p < 5\%$ ) und \*\* ( $p < 1\%$ ) gekennzeichnet.

Die Begründungen der Organisationen für Umfang und Rhythmus der Evaluationsverfahren lassen sich deutlich auf zwei Ebenen verorten. Die eine Gruppe an Organisationen führt vorwiegend die personellen Ressourcen als Argument für die Beschränkung von Umfang und Rhythmus an und ergänzt dies nur teilweise um Begründungen, die sich inhaltlich auf die Evaluation beziehen, wie etwa, dass man die Einschränkung auf x% aller Kurse als ausreichend erachte oder dass umfangreichere Evaluationen sowieso keinen Mehrgewinn erbringen würden (u.a. 269:3; 479:13). Die andere Gruppe an Organisationen führt inhaltliche Begründungen an, wobei teilweise auch die Perspektive der Teilnehmenden berücksichtigt wird. So will man beispielsweise schnelle Rückmeldungen gewährleisten, die Evaluation auf das Kerngeschäft beschränken, externe Vorgaben umsetzen und die Teilnehmenden und Lehrenden durch zu häufige, sich wiederholende Befragungen nicht überstrapazieren (u.a. 420:14; 428:19; 434:13; 669:23).

Bei langfristigen und bei mehrtätigen Angebotsformaten, wie etwa Berufsausbildungen, arbeitsmarktnahen Qualifizierungsmaßnahmen, Bildungsurlauben und Kursen zum Erwerb von Schulabschlüssen, führen die Organisationen mehrfach innerhalb eines Angebots Evaluationen durch. Dabei kommt die gesamte Palette an oben genannten Verfahren zum Einsatz. So werden beispielsweise nach jeder Unterrichtseinheit ein mündliches Feedback eingeholt, zweimal pro Jahr der Lernerfolg mithilfe von Beurteilungsbögen bewertet oder mindestens vierzehntägig Einzelgespräche geführt (297:24, 325:7; 434:21).

Über alle Organisationen hinweg betrachtet dominiert bezüglich des Zeitpunktes, wann innerhalb eines Bildungsangebots evaluiert wird, deutlich die Evaluation am Ende oder nach Abschluss des Angebots. Dieser Zeitpunkt wird nur von wenigen Organisationen begründet, so seien etwa die Teilnehmenden erst am Ende in der Lage ein Gesamtbild zu beurteilen (339:24). Nur wenige Organisationen geben an, dass sie ihre Angebote bereits zwischen Mitte und Ende der Veranstaltung evaluieren. Als Argument wird hierfür angeführt, dass man im Falle berechtigter Kritik noch intervenieren könne (292:7).

Zum Vorgehen bei der Datenerhebung (im Sinne der konkreten Datenbeschaffung während einer „Feldphase“) lassen sich in den Selbstreporten nur von 35 der 50 Organisationen Aussagen identifizieren, die zumeist sehr knapp ausfallen.<sup>70</sup> Sehr auffällig dabei ist, dass die Lehrenden bei der Datenerhebung eine Schlüsselrolle einnehmen: Bei Fragebogenerhebungen erhalten die Lehrenden die Bögen von der Organisation, sie vermitteln den Teilnehmenden den Sinn der Evaluation, übernehmen das Austeilen, Einsammeln, zum Teil auch das Zusammenfassen der Daten und die Übertragung in Formblätter und leiten die Bögen bzw. Ergebnisse an die Verantwortlichen weiter. Alternativ oder ergänzend zur Fragebogenerhebung fassen sie in Berichten das Kursgeschehen zusammen. Diese Aufgaben der Lehrenden werden von einigen Organisationen umfassend (u.a. 312:8), von anderen nur in Teilaspekten beschrieben (u.a. 276:3).

Einige wenige Einrichtungen weisen in ihrem Selbstreport ausdrücklich auf Maßnahmen zur Wahrung der Anonymität bei der Datenerhebung hin. So teilen zwar beispielsweise die Lehrenden Evaluationsbögen aus, verwenden auf Wunsch jedoch geschlossene Briefumschläge zum Einsammeln oder die Teilnehmenden übermitteln die Bögen mittels zentral bereitgestellter Behälter oder per Fax (u.a. 308:9).

---

70 Dies kann vermutlich auf zwei wesentliche Gründe zurückgeführt werden. Zum einen sieht das LQW-Modell keine Mindestanforderung zur Beschreibung der Datenerhebung vor, die Organisationen sind also nicht explizit verpflichtet, hierzu Angaben im Selbstreport zu machen. Zum anderen muss zusätzlich bedacht werden, dass es auch in der Methodenliteratur wenige Angaben zum konkreten Vorgehen bei der Datenerhebung gibt, so findet man beispielsweise kaum Informationen darüber, wie Fragebögen verteilt, eingesammelt und weitergeleitet werden, höchstens dazu, wer die Erhebung durchführen sollte und welche Effekte mit unterschiedlichen Erhebungspersonen verbunden sind (z.B. bei Diekmann, 2007, S. 466-470).



### 5.3 Datenauswertung und Datenbewertung

Bezüglich der Datenauswertung wurden die vorliegenden Selbstbeschreibungen der Evaluationspraxis nach Antworten auf die Frage „Wer wertet, wann, wie und mit welchen Hilfsmitteln die erhobenen Daten aus?“ durchsucht. Dabei wurde der Auswertungsbegriff sehr weit gefasst, da auch die Organisationen den Begriff sehr unterschiedlich breit verwenden: Zum einen wird Auswertung von den Organisationen sehr eng im Sinne einer Ergebniszusammenstellung, zum anderen sehr weit im Sinne einer umfassenden Analyse aufgefasst.

Zwar lassen sich bei 44 der 50 Organisationen Textstellen zur Datenauswertung identifizieren, allerdings hat keine der Organisationen alle oben gestellten Teilfragen (wer, wann, wie, welche Hilfsmittel) vollständig beantwortet und konkrete Informationen über das Vorgehen bei der Auswertung sind in den Selbstbeschreibungen rar gesät. Am häufigsten berichten die Organisationen über die Personen, welche die Datenauswertung durchführen. Es handelt sich überwiegend um das pädagogische Personal, das in der Rolle der Bereichsleitung oder der Projektverantwortlichen die Datenauswertung vornimmt, wobei sich häufig auch eine Datenanalyse anschließt. Bei dieser Analyse werden die Ergebnisse „durchgesehen“, „reflektiert“, „vergleichend analysiert“ und wichtige Ergebnisse „markiert“ (325:20; 418:17; 439:5; 721:27). Da die Lehrenden bereits bei der Datenerhebung eine so prominente Rolle einnehmen (siehe oben), verwundert es nicht, dass in mehreren Organisationen neben dem pädagogischen Personal auch die Lehrenden die Datenauswertung übernehmen und hierfür beispielsweise Testergebnisse zusammenstellen und Fragebogenergebnisse zusammentragen (u.a. 312:17; 669:48). Nur eine Organisation gibt dabei an, dass ihren Lehrenden für die Datenauswertung und Nachbereitung bezahlte Zeiten zur Verfügung stehen (211:77). In einigen Organisationen obliegt die Datenauswertung den Mitarbeitenden aus der Verwaltung, die fortlaufend oder semesterweise Daten eingeben und die aufbereiteten Ergebnisse an die Verantwortlichen weiterleiten.

Wie wird genau ausgewertet? Neben zahlreichen knappen Aussagen, dass man „qualitativ und quantitativ“ oder „statistisch“ auswerte, finden sich bei immerhin fünf Organisationen Angaben über eine Aggregation bei der Datenauswertung. Häufigkeiten und Mittelwerte werden pro Frage zunächst für den Kurs, dann für den Fachbereich und schließlich für die Gesamtorganisation zusammengefasst und bei einigen Organisationen über die Zeit hinweg vergleichend dargestellt, um Tendenzen ablesen zu können (u.a. 244:7). Die Organisationen setzen vorhandene Seminarver-

waltungsprogramme wie SQL-BASys, aber auch selbstprogrammierte Online-Tools und Excel-Dateien zur Unterstützung der Auswertung ein. Unter den wenigen Angaben über die Häufigkeit der Datenauswertung halten sich jährlich stattfindende Auswertungen die Waage mit häufigeren, z.B. semesterweisen Auswertungsfrequenzen.

Trotz der expliziten Anforderung im LQW-Modell, dass die Organisationen im Selbstreport nachweisen müssen, wie Analysen bewertet werden, und trotz der hohen Bedeutung, die dem Aspekt der Bewertung in der Literatur zur Evaluation beigemessen wird, lassen sich nur in gut der Hälfte der 50 Selbstreporte Angaben zum Vorgehen bei der Bewertung finden – und diese Angaben sind zumeist sehr knapp gehalten. Die Organisationen, die Angaben zur Datenbewertung gemacht haben, teilen sich in zwei Gruppen auf.

Bei der einen Gruppe an Organisationen bleibt das konkrete Vorgehen bei der Bewertung zumeist unklar und vage. Man erfährt lediglich wer bewertet, z.B. die Fachbereichsleitung oder in welchem Rahmen bewertet wird, z.B. bei der Studienkonferenz (491:14; 707:26). Nur einige Einrichtungen dieser Gruppe werden konkreter und beschreiben, dass sie qualitative Bewertungen vornehmen. Hierfür greifen sie auf einfache Schemata wie „Was läuft gut, was läuft schlecht?“ zurück oder setzen komplexere Bewertungsfragen ein: „Greifen die Konzepte? Sind Folgekurse geplant? Wird das Lernumfeld angenommen?“ (259:32; 333:15; 731:29) Man erfährt also im Grunde, welche Fragestellungen die Evaluationsauswertung leiten, doch wie die Daten bewertet werden, woran die Organisationen etwa erkennen, ob das Lernumfeld angenommen wird, bleibt meist verborgen.

Die andere Gruppe der Organisationen informiert in ihren Selbstreporten über klar definierte Sollwerte, die sich fast ausschließlich auf quantitative Daten beziehen. Als Sollwerte dienen zumeist Durchschnittsbewertungen, die man nicht unterschreiten will. Sollte dieser Fall dennoch eintreten, zieht man Konsequenzen. Neben Mittelwerten werden auch Anteilswerte zur Bewertung herangezogen, z.B. sollen 85% der Teilnehmenden Spaß am Kurs angegeben haben oder es sollen 80% den gesamten Kurs mit „gut“ oder „sehr gut“ bewerten – übrigens ein niedrig gestecktes Ziel, das leicht übertroffen wird, wie die Einrichtung festgestellt hat (276:21; 408:4). Unklar bleibt an mehreren Stellen, auf welcher Aggregationsebene die Sollwerte zum Tragen kommen: Reicht für eine negative Bewertung bereits eine Person aus, die den Sollwert unterschreitet? Reicht eine Unterschreitung der Durchschnittsnote bei einem Item aus? Oder muss stattdessen der Durchschnittswert des gesamten Kurses

zu niedrig ausfallen? Zwei Organisationen geben neben Sollwerten zudem auch an, die Vorjahresergebnisse bei der Bewertung zu berücksichtigen.

#### **5.4 Umgang mit Ergebnissen: Dokumentation, Kommunikation, Nutzung**

Die Analyse der Selbstreporte zeigt, dass die Dokumentation bezüglich der Evaluation von Weiterbildungsprozessen in den Organisationen auf zwei Ebenen erfolgt: Zum einen dokumentieren „Maßnahme-, Seminar- und Kursleiterberichte“ die Evaluationsergebnisse einzelner Bildungsmaßnahmen. Zum anderen wird auf einer darüber liegenden Ebene die Diskussion über die Evaluationsergebnisse einzelner und mehrerer Bildungsmaßnahmen und die abgeleiteten Konsequenzen dokumentiert, zumeist in Form von „Protokollen“. Beide Berichtsformen dienen einigen Organisationen zur Legitimation, zur Planung sowie zum Marketing und zur Information von Ausschüssen (u.a. 420:14; 483:8).

Bezüglich der Kommunikation über die Ergebnisse lassen sich im Wesentlichen drei Arten unterscheiden. Erstens berichten die Organisationen am häufigsten von der Rückmeldung der Ergebnisse an die Lehrenden, welche dezidiert im LQW-Modell gefordert wird. Diese Rückmeldung geschieht meist auf persönlichem Wege im Rahmen von Konferenzen oder Sitzungen, so etwa auf „Dozenten- und Fachbereichskonferenzen“, bei „Teamsitzungen“ oder während eines „Fortbildungstages“ (u.a. 269:13; 357:18; 501:25). Etwas seltener werden die Lehrenden auf persönlichem Wege durch die Fachbereichsleitenden in Einzelgesprächen informiert. Diese Form der Rückmeldung ist unabhängig vom Organisationstyp, dominiert jedoch deutlich in kleineren Einrichtungen mit weniger als fünf hauptamtlichen pädagogischen Mitarbeitenden. Mehrere Organisationen wählen auch den schriftlichen Weg, um ihre Lehrenden über die Evaluationsergebnisse zu informieren, und greifen auf Mail, Newsletter und Kursleiteranschreiben zurück. Beachtenswert ist, dass sich in einigen Organisationen die Lehrenden selbst über die Evaluationsergebnisse informieren (sollen), da sie ja bereits die ausgefüllten Fragebögen im Zuge der Datenerhebung gesehen oder gar selbst einen Seminarbericht verfasst hätten. Die Organisationen sehen die Lehrenden hier also in einer Selbstverantwortung zur Informationsbeschaffung (u.a. 428:29-31; 445:9,13).

Die zweithäufigste Form der Kommunikation über die Ergebnisse stellen Besprechungen und Konferenzen des leitenden Personals dar, z.B. die jährlichen Evaluationskonferenzen, die monatlichen Controllinggespräche oder die wöchentlichen Mit-

arbeiterrunden (u.a. 308:13; 439:7; 707:25). Hier steht nicht die Rückmeldung an die Lehrenden im Vordergrund, stattdessen tauschen sich die Mitarbeitenden über die Ergebnisse aus und es werden über Vorschläge der Lehrenden diskutiert und konkrete Maßnahmen beratschlagt. In diesen Besprechungen stehen häufig strukturelle und grundsätzliche Fragen auf der Tagesordnung. Neben den tendenziell eher internen Besprechungen finden auch Konferenzen statt, in denen die Ergebnisse dem Vorstand, dem Beirat oder anderen Gremien vorgestellt werden.

Als dritte Form der Kommunikation ist noch die Rückmeldung der Evaluationsergebnisse an die Teilnehmenden und die Information der Öffentlichkeit zu nennen. Dies geben allerdings nur drei Organisationen, allesamt Volkshochschulen, in ihrem Selbstreport explizit an. Während die eine Volkshochschule die Teilnehmendenvertretung im Rahmen einer Versammlung über die allgemeinen Ergebnisse informiert, veröffentlicht die andere die Ergebnisse im Programmheft. Die dritte Volkshochschule hat allgemeine Ergebnisse im Internet veröffentlicht, meldet aber zudem auch an Teilnehmende Ergebnisse und Konsequenzen zurück, sofern es sich um persönlich eingebrachte Kritik der Teilnehmenden handelte. (252:30; 479:10; 488:27)

Wie gehen die Organisationen praktisch mit den Evaluationsergebnissen um? Welche Konsequenzen ziehen die Organisationen also aus den Evaluationsergebnissen und wie nutzen sie die Ergebnisse? In erster Linie werden die Ergebnisse zur *Verbesserung, Optimierung und Entwicklung der Bildungsprozesse* eingesetzt – hierüber berichten 33 der 50 Organisationen. So passen die Organisationen insbesondere die Kurskonzeption sowie die Lehr- und Lernprozesse an, indem sie zu Tage getretene Bedürfnisse berücksichtigen, Inhalte austauschen, die Zielgruppendefinition überdenken oder konkrete pädagogische Interventionen vornehmen und zum Beispiel Stützunterricht bei Auszubildenden anbieten (u.a. 259:15; 491:23; 492:27). Daneben wird auch sehr häufig die Lerninfrastruktur mithilfe der durchgeführten Evaluationen optimiert, indem Heizungsanlagen früher eingeschaltet, EDV-Ausstattungen aufgerüstet oder die Reinigungszeiten angepasst werden (u.a. 269:11; 276:16; 491:15). Verbesserungen erzielen die Organisationen zudem beim Service und im Programmheft, indem Wartezeiten für Teilnehmende reduziert und die Qualität der Ausschreibungen und die Übersichtlichkeit im Programmheft gesteigert werden (u.a. 276:12; 292:8).

Neben der Nutzung der Evaluationsergebnisse zur Verbesserung und Weiterentwicklung nimmt bei den Organisationen die Nutzung für die *Bedarfsermittlung und Programmplanung* eine bedeutende Rolle ein: Die Evaluationsergebnisse fließen auf

den Planungskonferenzen ein, führen zu Themenverschiebungen, zum Ausbau spezieller Angebotsformen und zur Ausweitung der Angebote für spezielle Zielgruppen (u.a. 479:84; 483:12; 488:13).

Als Konsequenz aus den Evaluationsergebnissen führen viele Einrichtungen *Gespräche*, die vorwiegend von verantwortlichen pädagogischen Mitarbeitenden mit Lehrenden, deutlich seltener mit Teilnehmenden durchgeführt werden. Anlass für die Gespräche sind in der Regel schlechte Bewertungen, vereinzelt aber auch Unsicherheit oder Irritation mit Ergebnissen. Das Ziel der Gespräche besteht darin, Lösungen für die aufgetretenen Probleme zu finden, häufig werden jedoch zunächst die genauen Umstände eruiert, beispielsweise die konkreten Gründe für quantitative, negative Bewertungen erfragt, da die Zahlen in diesem Falle nicht für sich sprechen (u.a. 429:10).

Etwa jede fünfte Organisation nimmt die Ergebnisse ihrer Evaluationen als Anlass für *Fortbildungen und Qualifizierungen*, und zwar sowohl der internen Mitarbeitenden als auch der freiberuflichen Lehrenden. Beispielsweise werden für die internen Mitarbeitenden Fortbildungen zu Methodik/Didaktik oder zur Verfassung von Kursbeschreibungen initiiert (339:49; 418:15). Den externen Lehrenden werden Schulungsangebote vorgeschlagen oder sie werden zur kollegialen Beratung und anderen Qualifizierungsformen verpflichtet (429:10; 473:20). Hospitationen des hauptamtlich pädagogischen Personals werden dabei von einer Organisation als Überprüfung der Kursleitenden und von einer anderen als Unterstützungsangebot verstanden (308:13; 312:25).

Ebenfalls jede fünfte Organisation trifft anhand der Evaluationsergebnisse *Entscheidungen über die Weiterbeschäftigung von Lehrenden*, wenn diese bei der Evaluation schlecht abschneiden. Dabei handelt es sich zum einen um die gleichen Organisationen, die ihren Lehrenden zunächst auch Unterstützungsangebote wie Fortbildung und Qualifizierung anbieten, doch der größere Teil der Einrichtungen, die sich gegebenenfalls von Lehrenden trennen, sieht kein derartiges „Auffangnetz“ für ihre Lehrenden vor.

Neben den bisher genannten Konsequenzen und Nutzungsformen der Ergebnisse sind noch einige Einzelnennungen zu berichten. Hierzu zählen unter anderem die Erfolgskontrolle und Zielüberprüfung, die Weiterentwicklung der Preisgestaltung und der Geschäftsbedingungen sowie die Außendarstellung und Werbung, aber auch organisationsübergreifende Qualitätsentwicklung und die Auswahl von Tagungs-

stätten. Nicht zu unterschätzen ist darüber hinaus, dass Evaluation auch Bestätigung, Lob und Anerkennung bringt (u.a. 303:33).

Wie genau die Einrichtungen zu Konsequenzen gelangen, wie sie also den Weg von den Evaluationsergebnissen zur Handlung gestalten, bleibt in vielen Fällen vage und unklar. Jedoch wird diese Black-Box an einigen Stellen in den Selbstreports durch knappe Beschreibungen von Prozessen und Zuständigkeiten erhellt. Beispielsweise werden bei Problemen, die durch die Evaluation zu Tage getreten sind, die höheren Hierarchieebenen erst einbezogen, wenn keine Lösung herbeigeführt werden konnte oder die Ergebnisse werden zunächst von den Servicemitarbeitenden hinsichtlich Infrastruktur und Service durchgesehen und anschließend von der Leitung bezüglich pädagogischer Belange (u.a. 297:33; 721:16). Die Zuständigkeiten bei der Bearbeitung von Evaluationsergebnissen ergeben sich hier also aus unterschiedlichen Evaluationsinhalten und -gegenständen. Auch in anderen Organisationen ist eine Auftrennung der Evaluationsergebnisse zu unterschiedlichen Inhalten und Gegenständen zu beobachten, etwa in die zwei Felder Infrastruktur, Räume und Außengelände einerseits und Kursorganisation andererseits (252:38-66). Eine Organisation berichtet über zahlreiche analytische Aspekte, die bei der Analyse der Evaluationsergebnisse Berücksichtigung finden: Mängelbehebung kurzfristig vs. mittel- oder langfristig möglich, Verantwortlichkeit für das Problem in der Organisation oder außerhalb, Entscheidung über Maßnahmen kann innerhalb eines Teilbereichs der Organisation oder nur für die gesamte Organisation getroffen werden (317:26).

In personeller Hinsicht lässt sich feststellen, dass die Lehrenden gemäß ihrer üblichen Tätigkeiten häufiger bei Konsequenzen bezüglich der Lehr-Lern-Prozesse und der Kurskonzeptionen beteiligt sind, während Konsequenzen bezüglich Programmplanung, Personaleinsatz und Finanzfragen den disponierend und koordinierend Tätigen sowie der Leitung vorbehalten sind.

Die häufigsten Nutzungsformen und Konsequenzen sind in Abb. 41 nach Organisationstypen aufgliedert zu sehen. Durchweg nutzen die meisten Organisationen die Ergebnisse vorwiegend zur Verbesserung, Optimierung und Entwicklung. Unter den sozial bzw. kulturell orientierten Organisationen geben überdurchschnittlich viele an, Lehrende gegebenenfalls nicht weiter zu beschäftigen. Unter den weltanschaulich gebundenen Organisationen finden sich sehr wenige Organisationen, die als Konsequenz ihrer Evaluationsbemühungen Gespräche führen. Bei den Volkshochschulen finden sich sehr viele Organisationen, welche die Ergebnisse zur Bedarfsermittlung und Programmplanung verwenden, während dies nur für wenige kom-

merziell orientierte bzw. wirtschaftsnahe Organisationen gilt. Allerdings sind keine dieser Auffälligkeiten bezüglich des Organisationstyps statistisch signifikant. Auch lassen sich zur Organisationsgröße keine Zusammenhänge herstellen.

**Abb. 41: Unterschiede bezüglich Konsequenzen und Nutzungsformen der Evaluationsergebnisse bei verschiedenen Organisationstypen**

	sozial/ kulturell	weltanschau- lich gebunden	Volkshoch- schulen	kommerziell/ wirtschaftsnahe	Gesamt
	n = 10	n = 9	n = 20	n = 11	n = 50
Verbesserung, Optimierung, Entwicklung	7 (70%)	6 (75%)	15 (75%)	6 (50%)	33 (68%)
Bedarfsermittlung, Programmplanung	5 (50%)	3 (38%)	<b>12 (60%)</b>	<b>2 (17%)</b>	22 (44%)
Gespräche	4 (40%)	<b>1 (13%)</b>	7 (35%)	6 (50%)	18 (36%)
Fortbildung, Qualifizierung	2 (20%)	2 (25%)	3 (15%)	3 (25%)	10 (20%)
Lehrende: keine Weiterbeschäftigung	<b>4 (40%)</b>	1 (13%)	3 (15%)	2 (17%)	10 (20%)

Angegeben ist die Anzahl an Organisationen, die über die jeweiligen Konsequenzen und Nutzungsformen berichtet hat. Die Prozentangaben beziehen sich auf die Anzahl an Organisationen des jeweiligen Typs.

**Fett** hervorgehoben sind Häufigkeiten, deren standardisiertes Residuum vom Betrag größer gleich 1 ist. In der Tabelle finden sich keine Häufigkeiten, die statistisch signifikant von den erwarteten Häufigkeiten abweichen.

## 5.5 Zusammenfassung und Diskussion des Status quo

Bei den Organisationen stehen als Evaluationszwecke die Zielüberprüfung und die Verbesserung im Vordergrund, wobei vorrangig die Erreichung der eigenen Ansprüche aus dem Leitbild und der Definition gelungen Lernens überprüft werden soll. Eine Lernendenorientierung in dem Sinne, dass Evaluation auch für die Teilnehmenden einen Nutzen wie etwa Lernunterstützung oder Lernreflexion erbringen, also selbst zum gelungenen Lernen beitragen kann, wird selten als Evaluationszweck benannt. Die in der Literatur häufig diskutierte Legitimations-Funktion von Evaluation wird nur von vereinzelt Organisationen beschrieben. Dies ist eventuell darauf zurückzuführen, dass Legitimation von den Organisationen nicht auf der Ebene der konkreten Bildungsprozesse, sondern auf der Ebene der Gesamtorganisation, z.B. in Form des LQW-Testats verortet wird. Es wird nur selten von einzelnen, punktuellen Fragestellungen berichtet, die im Rahmen der Evaluation einmalig und fokussiert verfolgt werden. Stattdessen hat die Evaluation von Bildungsprozessen eher den Charakter eines kontinuierlichen Monitorings wie es im Rahmen der Evaluation etwa von Owen (2006) oder Patton (2008) beschrieben wird. Dieser Monitoring-Charakter steht auch im Einklang mit dem Ergebnis, dass überhaupt keine Einrichtung von einer prognostischen Evaluation berichtet, nach der eine Bildungsmaßnahme vor ihrer Durchführung evaluiert würde. Auf eine solche Evaluation ist das LQW-Modell im Qualitätsbereich 5 zur Evaluation der Bildungsprozesse nicht

ausgelegt, sondern dies ließe sich eher dem Qualitätsbereich 2 zur Bedarfserschließung zuordnen.

Die evaluierten Gegenstände lassen sich im Wesentlichen auf den in Kapitel 2 vorgestellten vier Dimensionen „Lernen“, „Lehren“, „Lerninfrastruktur“ und „Lernorganisation“ verorten. Dabei stehen bezüglich des Lernens der Lernerfolg, bezüglich des Lehrens die Lehrenden und ihr Unterricht, bezüglich der Lerninfrastruktur die Infrastruktur im Allgemeinen und die Räume sowie bezüglich der Lernorganisation der Service und die Mitarbeitenden im Vordergrund. Auch die Zufriedenheit im Ganzen im Sinne einer „Overall-Satisfaction“ als auch die Zufriedenheit mit einem konkreten Gegenstand wie etwa der Lerninfrastruktur nehmen eine prominente Rolle unter den Evaluationsinhalten ein. Die fünfte, übergeordnete Dimension „Systemumfeld“ wird im Rahmen der Evaluation der Bildungsprozesse nicht angesprochen. Auffällig ist, dass das Verständnis der Einrichtungen von einem Bildungsprozess nicht ausgewiesen wird und der evaluierte Bildungsprozess in seiner Ganzheit nie beschrieben wird. Die Organisationen fassen (möglicherweise in der Folge nicht klar beschriebener Bildungsprozesse) die Evaluationsgegenstände sehr kleinteilig auf, wobei die Grenze zu Bewertungskriterien verschwimmt, wie bereits in vorherigen Studien festgestellt werden konnte (vgl. Abschnitt 2.4.4).

Als Evaluationsverfahren werden am häufigsten Fragebögen für die Teilnehmenden und Lehrenden sowie Gespräche eingesetzt. Zu den Gesprächen zählen sowohl gezielt geplante Gespräche, z.B. Auswertungsgespräche zwischen Lehrenden und den verantwortlichen Mitarbeitenden, als auch Gesprächsformen im Seminarprozess und informelle Gespräche. Gerade die informellen Gespräche spielen für immerhin ein Drittel der Organisationen eine wichtige Rolle und ihr Nutzen wird mehrfach ausführlich begründet. Berichte nehmen im Rahmen der Evaluation eine Doppelfunktion ein, indem sie einerseits als Datenerhebungsverfahren dienen und andererseits die Evaluationsergebnisse dokumentieren.

Die meisten Organisationen wählen Strategien zur Konzentration der Evaluationsbemühungen, vor allem beschränkt man sich auf einzelne Programm- oder Themenbereiche und auf Veranstaltungen, die entweder mit neuen Lehrenden oder zu neuen Themen angeboten werden. Die Beschränkung wird einerseits mit Kapazitätsgründen, andererseits aus inhaltlichen Argumenten wie beispielsweise externen Vorgaben begründet. Eine Begründung in Bezug auf die Lernendenorientierung, dass Evaluation die Teilnehmenden und Lehrenden beispielsweise überstrapazieren kann, wird selten angeführt.



Als Zeitpunkt der Evaluation wählen die Organisationen zumeist das Ende ihrer Angebote, seltener die Mitte. Vor allem in langfristigen Kursen werden allerdings auch zusätzliche Evaluationen nach inhaltlich sinnvollen Abschnitten wie z.B. am Tagesende oder bei halbjährlicher Vergabe der Zwischenzeugnisse durchgeführt.

Bei der Datenerhebung übernehmen insbesondere die Lehrenden eine wichtige Rolle, da ihre Zuständigkeit teilweise bei der Verteilung der Fragebögen und der Erläuterung der Sinnhaftigkeit beginnt und bei der Zusammenfassung der Evaluationsergebnisse und dem Verfassen eines Seminarberichtes endet. Von der Integration und Verschränkung der Evaluation in den und mit dem Lehr-Lern-Prozess, wie es z.B. ausführlich in Beywl, Bestvater und Friedrich (2011) beschrieben wird, ist jedoch so gut wie gar nicht die Rede.

Es ist auffällig, dass die typischen Prozessschritte einer Evaluation, welche sich an die Datenerhebung anschließen und welche die Datenauswertung und -bewertung, den Umgang mit Ergebnissen, die Kommunikation über selbige, das Ziehen von Schlussfolgerungen und die Nutzung betreffen, in den Ausführungen der Organisationen häufig verschwimmen. Es ist daher nicht leicht möglich und auch nicht immer sinnvoll, klare Abgrenzungen der Schritte zu ziehen, da dies der komplexen Praxis nicht gerecht würde.

Auch bei der Datenauswertung und -bewertung verwenden die Organisationen zum Teil sehr unterschiedliche Bezeichnungen für ihre Vorgehensweisen und berichten unterschiedlich umfangreich darüber. Häufig fehlen zudem Informationen über den Zwischenschritt, wie die erhobenen Daten, beispielsweise die ausgefüllten Fragebögen, in Evaluationsergebnisse transformiert werden. Trotz dieser Abgrenzungsprobleme lässt sich feststellen, dass überwiegend das pädagogisch verantwortliche Personal die Auswertung vornimmt, teilweise unterstützt durch die Lehrenden und vereinzelt auch durch die Mitarbeitenden der Verwaltung. Dabei werden die quantitativen Daten unter anderem auf verschiedenen Organisationsebenen aggregiert und als Zeitreihen aufbereitet.

Trotz der Wichtigkeit, die der Bewertung von Daten im Rahmen von Evaluationen zugeschrieben wird und die sich auch in den Anforderungen des LQW-Modells niederschlägt, finden sich nur wenige konkrete Angaben darüber, wie diese Bewertung in den Organisationen vonstattengeht. Bei fast der Hälfte der Organisationen lassen sich in den Selbstreporten keine detaillierten Angaben zur Datenbewertung finden. Die Organisationen, die Angaben zur Datenbewertung liefern, unterteilen sich zum einen in Einrichtungen, die konkrete quantitative Sollwerte definiert haben, und

zum anderen in Einrichtungen, die keine Sollwerte berichten, jedoch beschreiben, von wem und in welchem Rahmen die Bewertung durchgeführt wird – nämlich zu meist von leitenden Angestellten im Rahmen von Konferenzen und Auswertungsge sprächen.

Die Weiterbildungsorganisationen dokumentieren einerseits die Evaluationsergeb nisse selbst und andererseits die daraus abgeleiteten Konsequenzen. Entsprechend der LQW-Anforderung, dass die Lehrenden über die Ergebnisse informiert werden müssen, sind hierzu zahlreiche Beschreibungen vorhanden, wobei auffällig viele Or ganisationen die Selbstverantwortung der Lehrenden betonen, sich eigenständig über die Evaluationsergebnisse zu informieren. Die Teilnehmenden und die Öffent lichkeit werden kaum über die Ergebnisse informiert und dementsprechend spielt auch die Nutzung der Ergebnisse zu Werbezwecken nur eine randständige Rolle. Stattdessen werden die Ergebnisse der Evaluationsbemühungen vorwiegend zur Verbesserung und Entwicklung sowie zur Bedarfsermittlung und Programmplanung verwendet. Vergleicht man die beschriebene Nutzung der Ergebnisse mit den be schriebenen Evaluationszwecken, lassen sich einige Unterschiede feststellen. Wäh rend zahlreiche Einrichtungen als Zweck ihrer Evaluationsbemühungen die Erfolgs kontrolle und Zielüberprüfung betonen, finden sich hierzu nahezu keine Nennungen bei den Konsequenzen und der Nutzung. Diese Schieflage ließe sich so erklären, dass die Erfolgs- und Zielkontrolle zwar durchgeführt wird, aber nur die daraus resultie renden direkten Verbesserungen in den Selbstreporten beschrieben werden. Eine zweite Schieflage besteht bezüglich der Bedarfsermittlung und Programmplanung. Diese wird nur sehr selten als Evaluationszweck benannt, wohl aber werden die Evaluationsergebnisse hierfür sehr häufig verwendet. Vermutlich ist die Konzeption der Evaluationen nicht so sehr auf die Programmplanung fokussiert, die Ergebnisse lassen sich aber dennoch sehr gut dafür einsetzen.

Die Analyse der Selbstreporte hat einige Zusammenhänge zwischen der Evaluati onspraxis und dem Organisationstyp hervorgebracht. So finden sich im Vergleich zu anderen Organisationen unter den *sozial bzw. kulturell orientierten Organisationen* weniger Einrichtungen, die angeben, Evaluation zum Zwecke der Verbesserung und Wissensgenerierung zu betreiben. Überdurchschnittlich viele dieser Organisationen verwenden Gespräche im Seminarprozess als Evaluationsverfahren und ziehen als Konsequenz der Evaluationen in Betracht, Lehrende nicht weiter zu beschäftigen. Unter den *weltanschaulich gebundenen Organisationen* finden sich vergleichsweise wenige, die als Evaluationsgegenstand das Lernen und Lehren benennen, die infor-

melle Gespräche führen und die als Konsequenz der Evaluationsergebnisse Gespräche führen. Die *Volkshochschulen* setzen häufiger als andere Organisationen Fragebögen für Lehrende sowie informelle Gespräche ein, dafür kommen deutlich seltener Gespräche im Seminarprozess zum Einsatz. Alle untersuchten Volkshochschulen beschränken sich bei der Evaluation auf eine Auswahl ihrer Angebote. Unter ihnen finden sich viele Organisationen, welche die Ergebnisse der Evaluation zur Bedarfsermittlung und Programmplanung verwenden. Die *kommerziell orientierten bzw. wirtschaftsnahen Organisationen* wollen sehr viel mit der Evaluation der Bildungsprozesse erreichen, denn sie benennen alle vier Hauptzwecke der Evaluation häufiger als andere –insbesondere die Verbesserung und Wissensgenerierung. Sie setzen häufiger als andere Organisationen Berichte als Evaluationsverfahren ein und verwenden die Ergebnisse seltener für die Bedarfsermittlung und Programmplanung. Insgesamt betrachtet lässt sich jedoch festhalten, dass es sich nur um ein paar wenige Unterschiede handelt, die zum Teil auch mit dem Zufall der Stichprobensatzung erklärt werden könnten. Die vier Organisationstypen unterscheiden sich folglich bezüglich einzelner Details ihrer Evaluationspraxis, jedoch lässt sich für keinen der vier Organisationstypen eine eigenständige und charakteristische Evaluationspraxis konstatieren. Auch die Organisationsgröße stellte sich bei der Analyse als kein relevantes Unterscheidungskriterium für die Evaluationspraxis heraus; der einzige gefundene Zusammenhang, dass Berichte häufiger in großen Organisationen eingesetzt werden, lässt sich auf den übermäßigen Anteil von kommerziell orientierten bzw. wirtschaftsnahen Organisationen unter den großen Einrichtungen zurückführen.

In der Gesamtschau lässt sich für die Evaluationspraxis in den LQW-Organisationen festhalten, dass es sich nahezu ausschließlich um interne bzw. Selbstevaluation handelt. Teilweise wird zwar Unterstützung von außen durch Diplomanden, Hochschulen oder Personen mit Kenntnissen empirischer (Forschungs-)Methoden in Anspruch genommen, aber keine der Organisationen hat eine Dienstleistungsfirma mit der Durchführung standardisierter Befragungen oder anderer Evaluationstechniken beauftragt. In keiner der analysierten Selbstbeschreibungen zur Evaluationspraxis sind Rückbezüge oder auch nur Hinweise auf Evaluationsansätze oder -modelle, wie sie in Abschnitt 2.6 berichtet wurden, zu finden. Im Kapitel 2 wurden drei Definitionen von „Evaluation in der Weiterbildung“ vorgestellt. Zunächst verwundert es nicht, dass die LQW-eigene Definition von Zech (2006) die von den Organisationen beschriebene Praxis am treffendsten widerspiegelt, zumal die Definition die Praxis anleiten soll. Doch ist auffällig, dass die Definition von Gerl (1986), der die didakti-

sche Funktion von Evaluation betont, nicht kompatibel mit der herausgearbeiteten Praxis ist, da nur wenige Organisationen überhaupt auf den Aspekt der Unterstützung der Lernenden eingehen. Auch die Definition von Reischmann (2006) erscheint nur zum Teil und nur für einzelne Organisationen anschlussfähig: Während die Datenerhebung zwar überwiegend systematisch erfolgt, verbleibt die von Reischmann geforderte begründete Bewertung der Daten bei vielen Beschreibungen wenig konkret.

### **5.6 Quintessenz für die Evaluationspraxis: Ein Fragenkatalog für die Reflexion und Gestaltung der Evaluation von Weiterbildungsprozessen**

Auf Basis der vorgestellten Ergebnisse zur durchgeführten Analyse der Selbstreports wurde eine Liste mit Fragen erstellt. Diese Fragen berücksichtigen die ausgewertete Fülle an Informationen zu den einzelnen Aspekten des Evaluationsprozesses aus den 50 Selbstreports und ermöglichen Bildungseinrichtungen (nicht nur solchen, die das LQW-Modell anwenden) eine Reflexion und zielgerichtete Gestaltung ihrer Evaluationspraxis. Die Liste kann zudem bei der Erstellung eines LQW-Selbstreports hilfreich sein, weil sie Anregungen enthält, über welche Aspekte der Evaluationspraxis konkrete Aussagen im Selbstreport getroffen werden können. Natürlich sollen diese Fragen nicht so verstanden werden, dass ein Selbstreport über alle diese Fragen Auskunft geben soll oder gar muss – dies würde dem Anspruch von LQW entgegenstehen, nicht normativ zu sein (vgl. Zech et al., 2006). Die Liste hat vielmehr einen unterstützenden Charakter, ähnlich wie die Arbeitshilfen zu den einzelnen Qualitätsbereichen von LQW, die unter [www.artset-lqw.de](http://www.artset-lqw.de) kostenlos im Internet verfügbar sind. Die Gliederung der Liste entspricht den oben verwendeten Zwischenüberschriften und folgt damit dem für die Analyse verwendeten Kategoriensystem, das einen typischen Evaluationsprozess abbildet. Da die Liste Reflexionsprozesse unterstützen soll, wurden die Fragen aus der „Wir“-Perspektive formuliert.

#### **Evaluationszwecke und Evaluationsgegenstände**

- Was bezwecken wir mit unserer Evaluation? Inwieweit wollen und können wir den Erfolg unserer Arbeit und die Zielerreichung kontrollieren, einzelne Bereiche und Aspekte verbessern und weiterentwickeln, Wissen und Erkenntnisse generieren und die Lehrenden und Lernenden unterstützen?

- Welchen Nutzen können und sollen unsere Evaluationsverfahren den Lernenden bringen?
- Welche Fragen soll die Evaluation beantworten?
- Wie definieren wir für uns den Bildungsprozess, den wir evaluieren? Welche Aspekte und Schritte umfasst dieser Bildungsprozess?
- Inwieweit soll die Evaluation der Öffentlichkeitsarbeit, der Legitimation, der Programmplanung, der Reflexion sowie der mikro- und makrodidaktischen Steuerung dienen?
- Welche weiteren Zwecke sind von Bedeutung?
- Über welche Aspekte des Lernens (z.B. Lernerfolg, Lernklima, Lerntempo) und des Lehrens (z.B. Methodik, Didaktik, Inhalte), der Lerninfrastruktur (z.B. Ausstattung, Umgebung, Lernmittel) und der Lernorganisation (z.B. Beratung und Information; Anmeldung; Freundlichkeit; Lernzeiten; Herkunft, Zusammensetzung und Anzahl der Teilnehmenden) soll die Evaluation Informationen liefern? Welche der Aspekte können und wollen wir überhaupt beeinflussen und verändern?
- Welche Bedeutung hat für uns die Erhebung von Zufriedenheit der Teilnehmenden im Rahmen der Evaluation, sowohl der Gesamt-Zufriedenheit als auch der Zufriedenheit mit Teilaspekten?
- Inwieweit sind geographische Informationen hilfreich, z.B. bezüglich der Kursauslastung unterschiedlicher Kursorte oder bezüglich der erreichten Teilnehmenden?
- Welche Gegenstände und Fragen der Evaluation leiten sich aus unserem Leitbild und unserer Definition gelungenen Lernens ab?

### **Evaluationsverfahren, Evaluationsdesigns und Datenerhebung**

- Welche Erhebungsverfahren können wir jenseits von Fragebögen einsetzen, z.B. Interviews, Beobachtungen, Materialanalysen?
- Auf welche vorhandenen Daten können wir bereits zurückgreifen? An welchen „indirekten“ Indikatoren können wir zudem den Erfolg unserer Arbeit erkennen, wie z.B. an Weiterempfehlungen und Wartelisten?
- Für welche unserer Teilnehmenden eignen sich Fragebögen?

- Welche Vergleichbarkeit der Daten ist sinnvoll und notwendig? Welche Themenblöcke in Fragebögen sind für alle Teilnehmenden relevant, welche nur für spezielle Zielgruppen oder Veranstaltungsformen?
- Inwieweit sollen offene Fragen ohne vorgegebene Antwortmöglichkeiten in Fragebögen gestellt werden?
- Welche Rolle spielen interne bzw. externe Tests und Prüfungen für uns und wie können diese bzw. ihre Ergebnisse für die Evaluation fruchtbar gemacht werden?
- Welche Rolle spielen Gespräche und gesprächsaktivierende Verfahren für uns bei der Evaluation? Mit wem wollen wir gegebenenfalls für die Evaluation Gespräche führen: mit Teilnehmenden, mit Lehrenden, mit Mitarbeitenden?
- Wie gestalten wir den Umfang unserer Evaluation? Streben wir eine Vollerhebung an oder beschränken wir uns gezielt auf z.B. einzelne Themenbereiche, auf bestimmte Veranstaltungsformen oder Angebote mit einer festgelegten Mindestdauer, auf zuvor schlecht bewertete Veranstaltungen oder auf Veranstaltungen mit neuen Lehrenden, Themen, Kunden oder an neuen Orten? Wie viele Daten können wir bei der späteren Auswertung bewältigen?
- Wie oft innerhalb eines Bildungsangebotes und zu welchen Zeitpunkten führen wir Evaluationen durch: einmal oder mehrmals; zu Beginn, in der Mitte, am Ende, danach oder zu einem anderen Zeitpunkt?
- Wie vermeiden wir die Überstrapazierung von Teilnehmenden bei der Datenerhebung – sowohl bezüglich Häufigkeit als auch der Form der Evaluation?
- Welche Rolle übernehmen die Lehrenden bei der Datenerhebung?
- Produzieren unsere Verfahren die Daten, die wir mit Blick auf unsere Evaluationszwecke benötigen? Liefern uns die Instrumente also z.B. tatsächlich Verbesserungsvorschläge, und informieren sie uns darüber, inwieweit wir unsere Ziele erreicht haben und erfolgreich waren?

### **Datenauswertung und Datenbewertung**

- Wie werten wir unsere Daten aus? Inwieweit und auf welchen Ebenen werden die Daten aggregiert: auf Ebene der einzelnen Veranstaltungen, der Veranstaltungsformen, der Themen- oder Fachbereiche, der Organisationsbereiche, der Gesamtorganisation?

- Welche Berechnungen und Zusammenstellungen werden bei den quantitativen Daten benötigt und beherrscht?
- Wie bereiten wir die qualitativen Daten auf?
- Welche Hilfsmittel kommen bei der Datenauswertung zum Einsatz (z.B. Computerprogramme) und welche Ressourcen werden dafür benötigt?
- Wer wertet die Daten aus? Welche Unterstützung können die Lehrenden und die Mitarbeitenden der Verwaltung geben, welche benötigen sie? Welche Unterstützung von außen benötigen wir und welche können und wollen wir uns leisten?
- Wie bewerten wir qualitative und wie die quantitativen Daten? Welche Leitfragen, Kriterien, Sollwerte sind dabei von Bedeutung? Welche Vergleichswerte ziehen wir heran, z.B. Vorjahresergebnisse und/oder die Ergebnisse unterschiedlicher Angebote oder Fachbereiche?
- Wie viele negative Rückmeldungen, gleich ob qualitativ oder quantitativ, sind wir bereit zu akzeptieren? Gibt es Bereiche oder Gegenstände, in bzw. zu denen wir keine negativen Ergebnisse dulden?

#### **Umgang mit Ergebnissen: Dokumentation, Kommunikation, Nutzung**

- Wie dokumentieren wir die Evaluationsergebnisse und die Diskussion über selbige? Wie umfangreich muss diese Dokumentation sein? Wann muss wer auf die Ergebnisse zugreifen können?
- Wer wird von den Ergebnissen zu welchem Zweck und auf welche Art informiert: die Teilnehmenden, die Lehrenden, die Mitarbeitenden, die Financiers, die Öffentlichkeit?
- Zu welchen Themen (z.B. Lernorganisation, Lerninfrastruktur, Lehrende, Lernen, ...) und in welchen Arbeitsbereichen nutzen wir die Ergebnisse und wer ist jeweils dafür zuständig?
- Wozu nutzen wir die Ergebnisse: zur Verbesserung und Weiterentwicklung, zur Bedarfsermittlung und Programmplanung, als Anlass zu Gesprächen oder Fortbildungen und Qualifizierungen, als Personalentscheidungshilfe? Inwieweit nutzen wir die positiven Ergebnisse zur Mitarbeitermotivierung und Verteilung von Lob?

- Welche Konsequenzen ergeben sich aus den Ergebnissen zu den vorgenannten Bereichen? Legen die Ergebnisse z.B. nahe, Angebote für bestimmte Zielgruppen oder Regionen auszubauen oder einzustellen?
- Inwieweit nutzen wir die Ergebnisse so, wie ursprünglich bezweckt?
- Wie ziehen wir Konsequenzen? Inwieweit berücksichtigen wir dabei, a) wie gravierend festgestellte Mängel für die Betroffenen und Beteiligten sind, b) wie schnell sich die Mängel beheben lassen, c) in welchem Verhältnis Aufwand und Nutzen der Mängelbeseitigung stehen, d) inwieweit die Verantwortung in unserer Organisation liegt, e) wer in unserer Organisation bei der Umsetzung von Konsequenzen einbezogen werden sollte?
- Falls als Konsequenz der Ergebnisse Gespräche geführt werden: Wer führt diese mit wem und mit welchem Ziel?
- Welche Unterstützungsangebote können und wollen wir Lehrenden mit schlechten Bewertungen anbieten? Nach welchen Kriterien entscheiden wir, uns von Lehrenden zu trennen?
- Welche Verschränkungen der Evaluationsanstrengungen mit dem Controlling, der Bedarfserschließung und dem Lehr-Lern-Prozess sind möglich und sinnvoll?



## **6 Herausforderungen in der Evaluationspraxis**

Das sechste Kapitel widmet sich den Herausforderungen, denen Weiterbildungsorganisationen in Bezug auf die Evaluation ihrer Bildungsprozesse gegenüber stehen. Kernstück des Kapitels bilden deshalb die Ergebnisse von 24 Telefoninterviews mit Personen in Weiterbildungsorganisationen, die unmittelbar am Evaluationsprozess in ihren Einrichtungen beteiligt sind. Diese Personen werden im Folgenden zumeist als Organisationsvertreter/innen und nicht als Befragte bezeichnet, um deutlich zu machen, dass es nicht um ihre Person geht, sondern um sie als Informationslieferanten über die Evaluationspraxis in den jeweiligen Organisationen. Bevor im Detail die Interviewergebnisse vorgestellt werden, beginnt Kapitel 6 im Abschnitt 6.1 mit einer kurzen Darstellung der Herausforderungen, die sich in den Selbstreporten zeigen. Als Einstieg in das Datenmaterial der Interviews werden dann im Abschnitt 6.2 zunächst systematisch gegliederte Zusammenfassungen der Inhalte von vier Interviews vorgestellt. Es schließt sich in den Abschnitten 6.3 und 6.4 eine systematische Darstellung der Herausforderungen an, die im Zusammenspiel mit verschiedenen Faktoren wie dem Organisationstyp und der Organisationsgröße diskutiert und im Unterkapitel 6.5 durch weitere relevante Aspekte, wie etwa die geplanten Änderungen der Evaluationspraxis und dem Evaluationsverständnis, ergänzt werden. Den Abschluss des Kapitels bildet die Vorstellung einer Typologie der Herausforderungen, welche die Interviews aufgrund der in ihnen fokussierten Herausforderungen vier verschiedenen Typen zuordnet (6.6).

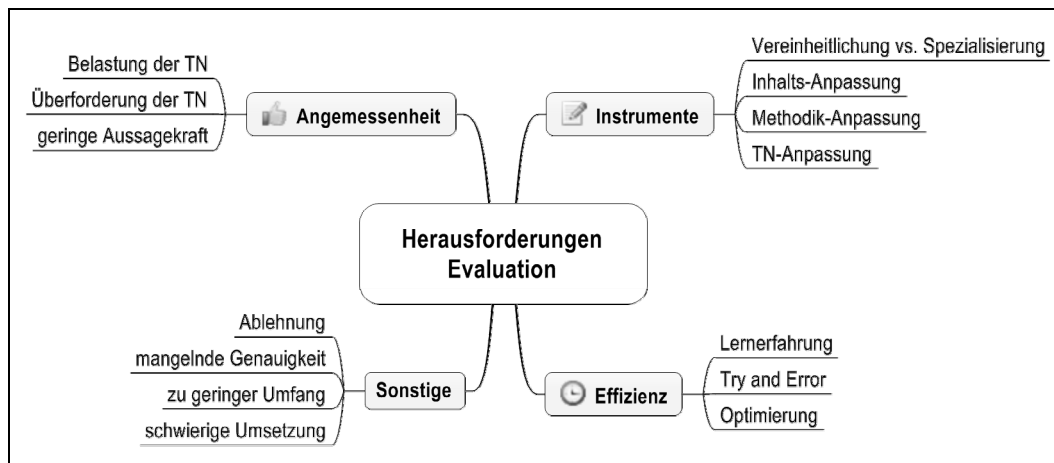
### **6.1 Von Organisationen berichtete Herausforderungen in den Selbstreporten**

Insgesamt haben 41 der 50 Organisationen Probleme und Herausforderungen in ihrem Selbstreport beschrieben. Die Abb. 42 präsentiert die in den Selbstreporten identifizierten und kategorisierten Problemfelder als Mindmap. Im Folgenden werden die drei Hauptproblembereiche „Instrumente“, „Effizienz“ und „Angemessenheit“ sowie weitere Einzelnennungen, die in der Mindmap auf dem Ast „Sonstiges“ verzeichnet sind, kurz erläutert.<sup>71</sup>

---

<sup>71</sup> Eine detaillierte Darstellung der Herausforderungen findet sich in Rädiker (2012).

Abb. 42: Übersicht über die identifizierten Problemfelder (Quelle: Rädiker, 2012, S. 48)



Mit großem Abstand vornan stehen Probleme, welche die Instrumente betreffen und sich vier Problembereichen zuordnen lassen: Erstens benennen die Organisationen das Spannungsfeld zwischen einerseits der Vereinheitlichung zwecks Vergleichsmöglichkeit der Ergebnisse und andererseits dem Spezialisierungsbedarf der Instrumente für bestimmte Bildungsinhalte und Kursformen. Zweitens stehen die Organisationen vor der Aufgabe, ihre Instrumente inhaltlich überarbeiten zu müssen, vor allem um den Inhalt an die Definition gelungenen Lernens und das Leitbild anzupassen. Drittens stehen für einige Organisationen Änderungen bezüglich der Methodik an, indem beispielsweise unausgewogene Antwortskalen in Fragebögen optimiert, offene Fragen in Fragebögen neu integriert und Internetbefragungen eingeführt wurden. Und viertens geben mehrere Organisationen an, dass sie ganze Instrumente für spezielle Teilnehmendengruppen, wie etwa die Teilnehmenden von „Deutsch als Fremdsprache“-Kursen, anpassen oder neu entwickeln mussten, weil die Standard-Papierfragebögen nicht verstanden wurden und teilweise auch nicht zielgruppenadäquat waren.

Bezüglich des zweithäufigsten Problembereichs, der Effizienz, lässt sich eine als klassisch zu bezeichnende Lernerfahrung erkennen, die man deutlich dem Bereich „Erfahrungslernen“ zuordnen kann und deren Aspekte von den Organisationen in unterschiedlichem Maße hervorgehoben werden: Die Organisationen beginnen ihre Evaluationsanstrengungen damit, umfangreich, häufig alle Kurse, zu evaluieren und produzieren eine große Menge an Daten ohne deren genaue Nutzung zu antizipieren. Später, vor allem bei der Auswertung stellt man fest, dass diese Datenmassen hohe Personal- und Zeitressourcen erfordern, zugleich jedoch die Erkenntnisse und deren Verwertbarkeit gering ausfallen. Nach mehrjährigem Ausprobieren und Op-

timieren der Evaluationsstrategie rückt die Frage nach dem Nutzen stärker in den Vordergrund und die Organisationen beschränken sich zunehmend auf ausgewählte Programmbereiche oder Veranstaltungen und reduzieren die Evaluationsfrequenz. Einfluss auf dieses Problem der Effizienz hat weniger der Organisationstyp, denn die Organisationsgröße: Je größer die Mitarbeitendenzahl ist, desto häufiger werden Probleme bezüglich der Effizienz genannt.

Als dritthäufigstes Problemfeld stellt sich die Angemessenheit der verwendeten Evaluationsverfahren heraus. Zum Teil die Frage nach der Effizienz fortführend, zum Teil unabhängig davon beschreiben die Organisationen in ihren Selbstreporten von einer geringen Aussagekraft der erhobenen Daten. Vor allem heben sie jedoch Abstumpfungstendenzen und Abwehr der Teilnehmenden hervor, die insbesondere dann auftreten können, wenn die von den Teilnehmenden geäußerte Kritik nicht bearbeitet wurde oder werden konnte.

Neben die großen Herausforderungen Instrumente, Effizienz und Angemessenheit der Verfahren kommen noch weitere Herausforderungen hinzu, die zwar tendenziell organisationsspezifischer Natur sind und keine großen erkennbaren Linien widerspiegeln, aber dennoch nicht unerwähnt bleiben sollen: Sie reichen von der Ablehnung und Hinterfragung der Sinnhaftigkeit der Evaluation seitens der Teilnehmenden und Lehrenden über zu geringen Rücklauf, also einem Problem der Datenerhebung, bis hin zu mangelnder Genauigkeit und Zuverlässigkeit der Evaluation (die keine gesicherten Schlussfolgerungen erlaubt) und hin zu Schwierigkeiten bei der Umsetzung der Ergebnisse.

## **6.2 Case Summarys vier ausgewählter Organisationen**

Die folgenden Fallzusammenfassungen stellen die Inhalte von vier Interviews in komprimierter und zugleich systematisierter Form dar, wobei sich die Darstellung nach den Hauptthemen des Interviewleitfadens gliedert. Unter den Punkten „Organisation“ und „Evaluationspraxis“ finden sich zudem kondensierte Informationen zum Hintergrund der Organisation und zur Evaluationspraxis aus dem begleitend eingesetzten Online-Fragebogen, wobei aus den Interviews der Evaluationszweck herausdestilliert und unter dem Punkt „Evaluationspraxis“ ergänzt wurde.

Es wurde von jedem der in Abschnitt 4.2.1 vorgestellten Organisationstypen (sozial/kulturell; weltanschaulich gebunden; Volkshochschule; kommerziell orientiert bzw. wirtschaftsnah) eine Organisation ausgewählt. Es wurde jeweils das zuerst mit

einem Organisationstyp geführte Interview ausgewählt, wobei es sich (zufälligerweise) um die vier zuerst geführten Interviews handelt, da hierunter alle vier Organisationstypen vertreten waren. Die Case Summaries geben den Inhalt der Interviews in kondensierter und gegliederter Form wieder. Sie verbleiben auf der beschreibenden Ebene und enthalten keine Wertungen oder Interpretationen. Die Case Summaries stellen den ersten analytischen Schritt dar und waren besonders hilfreich, da sie die jeweiligen Organisationen bezüglich der Aspekte, die für die Forschungsfragen relevant sind, porträtieren. Sie erlaubten es, die Herausforderungen in der Evaluationspraxis, welche die vier Organisationen zu meistern haben, im größeren Kontext zu betrachten. Während der Erstellung der Case Summaries wurden über die beschreibende Ebene hinausgehende Interpretationen, Hypothesen und Auffälligkeiten (z.B. Widersprüche) in Notizen festgehalten, um diese bei der anschließenden kategorienbasierten Auswertung einbeziehen zu können.

### 6.2.1 Sozial/kulturell orientierte Organisation

**Organisation** Es handelt sich um eine mittelgroße Heimvolkshochschule mit Beherbergungsbetrieb, in der neun hauptamtlich pädagogische Mitarbeitende arbeiten, die 400 Veranstaltungen im Jahr 2008 durchgeführt bzw. organisiert haben. Zielgruppe sind vor allem Schülerinnen und Auszubildende, Studierende, Lehrer/innen und Erzieher/innen, pädagogische Fachkräfte und die „interessierte Öffentlichkeit“. Schwerpunkte des Bildungsangebotes liegen bei der Vermittlung berufsfeldübergreifender Schlüsselkompetenzen, bei Lehrer- und Erzieherfortbildungen, bei der politischen Jugendbildung, bei der Umweltbildung, bei Familienseminaren sowie bei Theater und Kultur. Als hauptsächliche Veranstaltungsform bietet die Organisation (Halb-)Wochen- und Wochenendseminare sowie Akademieabende an.

**Evaluationspraxis** Die Organisation evaluiert folgende Aspekte: Methoden, Arbeitsprozess, Gruppenklima, Seminarleitung und Ziele. Als Evaluationsinstrumente kommen Abschlussrunden und Fragebögen zum Einsatz. Jede Veranstaltung wird mündlich evaluiert und alle zwei Jahre wird eine schriftliche Evaluation mit einer Stichprobe von 500 Teilnehmenden durchgeführt. An der Evaluation beteiligt sind die pädagogischen Mitarbeitenden sowie die Mitarbeitenden der Verwaltung, welche die Auswertung übernehmen. Die Einrichtung verfolgt mit der Evaluation den Zweck, qualitativ hochwertige Bildungsveranstaltungen anzubieten, so dass Teilnehmende wiederkommen. Die Evaluation dient der Heimvolkshochschule darüber hinaus der Qualitätssicherung und -entwicklung. Rückmeldungen während eines

Seminars helfen dabei einzuschätzen, ob man auf dem richtigen Weg war bzw. ist, und um das Seminar kurzfristig anzupassen.

**Herausforderungen bei der Evaluation** Die Teilnehmenden haben oft Schwierigkeiten, die abgefragten Seminarziele im Fragebogen zu verstehen und können mit den Aussagen nichts anfangen (z.B. „In der Veranstaltung habe ich Sichtweisen aus anderen Nationen und Kulturen kennengelernt und Möglichkeiten entwickelt mit kulturellen Unterschieden umzugehen.“). Bei der Analyse der Daten muss deshalb berücksichtigt werden, ob die Teilnehmenden die Fragen richtig verstanden haben und inwieweit sich vor diesem Hintergrund jeweils Aussagen treffen lassen. Darüber hinaus betont die interviewte Leitung das Problem, das richtige Maß zwischen fördern und fordern der Lehrenden (hauptsächlich interne pädagogische Mitarbeitende) zu finden. Problematisch ist zudem, dass unterschiedliche Teilnehmende unterschiedliche Wertungen vornehmen: Erzieher/innen werten generell besser als Lehrer/innen, während junge Erwachsene die Veranstaltungen der Heimvolkshochschule am kritischsten sehen.

**Geplante Änderungen der Evaluationspraxis** Stichproben sollen in Zukunft systematischer zusammengestellt werden. Interne Vergleiche von Lehrenden, unter denen einige positiv hervorstechen, sollen behutsam systematisiert werden. Eine Nachbefragung nach drei oder sechs Monaten wäre schön, ist aber nicht geplant.

**Unterstützungsbedarf** besteht bezüglich der Entwicklung eines bezahlbaren und guten Instrumentariums, das auch eine Nacherhebung beinhaltet.

**Nutzen und Nutzung der Evaluation** Die Evaluationsergebnisse werden direkt im Seminarprozess genutzt, um Änderungen am Ablauf oder Inhalt vorzunehmen, denn in der Organisation arbeitet man „sehr teilnehmerorientiert“ (I02:88)<sup>72</sup>. Eine Nutzenerhöhung der Evaluation könnte erreicht werden, indem die Langzeitwirkung ermittelt wird und indem die Ergebnisse besser für die Personalentwicklung bei schlechten Bewertungen der Lehrenden genutzt werden. Die Öffentlichkeit wird nicht über die Ergebnisse selbst informiert, sondern über die Tatsache, dass die Heimvolkshochschule überhaupt evaluiert.

**Benötigte Kompetenzen für die Evaluation** Um Bildungsprozesse effizient evaluieren zu können, benötigt man ein „gewisses Standing“ (I02:118), um als Seminarlei-

---

72 Die Zahlen in Klammern geben in diesem Kapitel die Quelle der Zitate an: Die Angabe vor dem Doppelpunkt informiert über die Nummer der/des Organisationsvertreters/in und die Zahl hinter dem Doppelpunkt steht für die Absatznummer des entsprechenden Transkripts.

tung die Rückmeldungen der Teilnehmenden auszuhalten. Ein „pädagogisches Gespür“ (I02:118) ist notwendig, um in einer Abschlussrunde dechiffrieren zu können, was die Teilnehmenden bei ihrem Feedback genau meinen. Man sollte zudem einen Fragebogen konstruieren können, der die Erreichung pädagogischer Ziele erhebt, sowie Evaluationsergebnisse in Maßnahmen und Änderungen umsetzen können, die Potenziale der lehrenden Mitarbeitenden entwickeln und freisetzen und diese nicht in Angst und Schrecken versetzen.

**Vorhandene Kompetenzen** Die benötigten Kompetenzen sind in der Heimvolkshochschule vorhanden, aber man lernt natürlich noch dazu. Erlangt wurden die Kompetenzen durch das Durchführen der Seminare und vor allem die Kommunikation darüber, so dass man inzwischen angstfrei über erweiterte Evaluation mithilfe von Fragebögen sprechen kann.

**Gewünschte Lerninhalte eines Seminars zur Evaluation** I02<sup>73</sup> möchte in einem Seminar die Frage behandeln, ob sich aus den Ergebnissen noch mehr ableiten lässt, als bisher berücksichtigt wurde. Zudem wäre interessant, mehr über den Beitrag der Evaluationsprozesse zur System- und Organisationsentwicklung zu erfahren.

**Verständnis von Evaluation** Evaluation heißt für I02 rückblickend auswerten, ob man die angestrebten Ziele erreicht hat: wie es einem auf dem Weg gegangen ist, was gut war und beibehalten werden sollte und wo man nachsteuern sollte. Andere Personen in der Organisation würden laut I02 den Kontrollaspekt von Evaluation stärker betonen.

### 6.2.2 Weltanschaulich gebundene Organisation

**Organisation** Es handelt sich um eine große Bildungseinrichtung einer Gewerkschaft mit 28 hauptamtlich pädagogischen Mitarbeitenden und einem Beherbergungsbetrieb. Die Organisationsvertreterin hat keine Angaben über die Anzahl durchgeführter Bildungsveranstaltungen 2008 gemacht. Zielgruppe sind vor allem Betriebsräte, Vertrauensleute, Jugendausbildungsvertreter, Sicherheitsbeauftragte, Gewerkschaftsmitglieder, ehrenamtliche und hauptamtliche Referenten/innen. Der Schwerpunkt der Bildungsangebote liegt im beruflichen Bereich und umfasst zwei Seminarreihen sowie gesellschaftspolitische, aufgabenbezogene und politische Bildung. Als hauptsächliche Veranstaltungsform bietet die Organisation Tagungen, Konferenzen, Wochenendseminare und Ein- bis Sieben-Tagesseminare an.

---

73 Das Kürzel „I02“ steht für den zweiten interviewten Organisationsvertreter.

**Evaluationspraxis** Die Bildungseinrichtung evaluiert folgende Aspekte: Inhalte, Methoden, Didaktik sowie Rahmenbedingungen (Kultur, Sport, Unterbringung, Service, Gastronomie). Als Evaluationsinstrumente kommen Fragebögen, vertiefende Interviews mit Teilnehmenden, Feedbackrunden in den Seminaren und Gespräche mit Beschäftigtengruppen zum Einsatz. In jedem Seminar werden Fragebögen ausgefüllt und ausgewertet, jährlich findet eine 14-tägige vertiefende Evaluation statt. Die Einrichtung verfolgt mit der Evaluation den Zweck, einen Blick für Mängel zu bekommen, um diese zu beseitigen, aber vor allem um Anregungen der Teilnehmenden offen aufzunehmen. Evaluation dient dazu, mithilfe von Rückmeldungen der Teilnehmenden den Inhalt und die Didaktik stärker in den Fokus zu nehmen und auf Nützlichkeit für die Teilnehmenden zu hinterfragen. Die Evaluation erlaubt es, die eigenen Vorstellungen, von dem, was gut ist, zu hinterfragen.

**Herausforderungen bei der Evaluation** Grundlegende Herausforderungen bestehen inzwischen keine mehr, nachdem in allen Bereichen Evaluation eingeführt wurde und auch ausgewertet wird. Allerdings mussten die Lehrenden erst überzeugt werden. Ob jetzt alle vollständig überzeugt sind, bezweifelt I03, geht aber von 90% sowohl bei den Lehrenden als auch anderen Beschäftigtengruppen aus. Problematisch ist die Routinebildung: Man reagiert nicht mehr so schnell wie am Anfang und leitet Hinweise verzögert weiter.

**Geplante Änderungen der Evaluationspraxis** In Zukunft sollen zusätzliche Themenbereiche in die Evaluation integriert und abgefragt werden, wie z.B. die Kinderbetreuung und der externe Service, um bei der Evaluation alle Beschäftigten zu berücksichtigen, die in irgendeiner Weise mit den Teilnehmenden zu tun haben.

**Unterstützungsbedarf** bei der Evaluation besteht keiner, denn die internen Mitarbeitenden bringen durch ihre Ausbildung Kenntnisse über Evaluation, statistische Auswertung und Fragebögen mit.

**Nutzen und Nutzung der Evaluation** Der Nutzen der Evaluation besteht in der Qualitätssicherung, der Entwicklungsmöglichkeit, der Erhebung von Bildungsnotwendigkeiten und von subjektiven Bildungsbedarfen. Die Durchführung von Workshops für alle Beschäftigten hilft dabei, Schnittmengen sichtbar zu machen, ins Gespräch zu kommen und Wertschätzung zu erhöhen. Evaluationsergebnisse werden zur Veränderung von Inhalten und Methoden genutzt und dienen als Entscheidungsgrundlage für Weiterbildungen der Mitarbeitenden. Die Ergebnisse sind ferner Gegenstand von Besprechungen, in denen um Veränderungen gerungen wird, häufig auch um Veränderungen von Haltungen. Der Nutzen der Evaluation könnte erhöht

werden, indem die festgestellten Probleme auch wirklich angegangen werden und indem man die nicht engagierten Personen dazu bringt, sich zu engagieren. Alle Mitarbeitenden werden über die Evaluationsergebnisse informiert, die Öffentlichkeit jedoch nicht.

**Benötigte Kompetenzen für die Evaluation** Um Bildungsprozesse effizient evaluieren zu können, benötigt man fachliche, inhaltliche Kompetenzen wie etwa zum Thema Arbeitsrecht. Zudem sind didaktische (nicht methodische) Kompetenzen notwendig, um Lernprozesse besser gestalten zu können – denn diesbezüglich „muss man schon verstehen, was die Menschen einem sagen wollen“ (I03:93).

**Vorhandene Kompetenzen** Die benötigten Kompetenzen sind mal mehr, mal weniger vorhanden. Die Mitarbeitenden bringen sie mit oder bilden sich entsprechend weiter.

**Gewünschter Lerninhalt eines Seminars zur Evaluation** I03 möchte lernen, mit Zeit effektiver umzugehen.

**Verständnis von Evaluation** Evaluation bedeutet für I03, eine Bestandsaufnahme und Untersuchung an einem Gegenstand vorzunehmen und sich den positiven und negativen oder kritischen Momenten einer Tätigkeit auszuliefern. Evaluation hat demnach hohe reflexive Momente. Die anderen Mitarbeitenden der Organisation hätten laut I03 ein unterschiedliches Verständnis von Evaluation und einige würden Evaluation auch als Mehrarbeit und lästig bezeichnen.

### 6.2.3 Volkshochschule

**Organisation** Es handelt sich um eine kleine Kreis-Volkshochschule (KVHS) mit vier hauptamtlich pädagogischen Mitarbeitenden, mehr als zehn Außenstellen und einem Volumen von 1.075 Veranstaltungen im Jahr 2008. Zielgruppe sind die Bewohner des Landkreises. Als Schwerpunkt der Bildungsangebote ist die KVHS bemüht, alle Bildungswünsche des Landkreises zu erfüllen. Vorwiegend werden Abendkurse angeboten.

**Evaluationspraxis** Die Organisation evaluiert folgende Aspekte: Zufriedenheit mit dem besuchten Kurs, den Lehrenden, dem Anmeldeverfahren und den räumlichen Gegebenheiten. Sie erfragt zudem, wie die Teilnehmenden von einem Kurs erfahren haben. Als Evaluationsinstrumente kommen einmal pro Jahr in ausgewählten Programmbereichen Papier-Fragebögen sowie zusätzlich Online-Fragebögen für EDV-Kurse zum Einsatz. An der Evaluation sind folgende Personen bzw. Personengrup-



pen beteiligt: Qualitätsbeauftragte (Organisation der Evaluation); Verwaltungskräfte (Druck und Verschickung Fragebögen); Außenstellenleitungen (Durchführung in der Veranstaltung); Honorarkraft (Auswertung). Die KVHS verfolgt mit der Evaluation den Zweck, die Qualität der Bildungsangebote zu verbessern und den Ansprüchen der Teilnehmenden gerecht zu werden.

**Herausforderungen bei der Evaluation** Freiwillig kommende Teilnehmende geben positive Bewertungen ab, weil sie sonst gar nicht gekommen wären, während unfreiwillig kommende Teilnehmende, wie etwa bei Integrationskursen, alles negativ beurteilen, egal wie und wo der Kurs ist. Die eigentliche Qualität der Veranstaltungen ist nur schwer zu erfassen. Seiteneffekte von Bildung, wie z.B. dass man den „Tischnachbarn“ im Seminar kennengelernt hat und die mit solchen Aspekten verbundene Emotionen, sind schwierig in verständliche Fragen für Erhebungen zu übertragen. Außerdem gibt es nur selten negative Äußerungen – wenn es etwas Negatives zu Lehrenden gibt, weiß man bereits, um wen es geht, und die Evaluation bringt daher keine neuen Erkenntnisse. Problematisch ist auch die Organisation der Evaluation, da die Volkshochschule sehr viele Außenstellen hat und eine breite Fläche bedient. Dabei muss die Qualitätsbeauftragte z.B. entscheiden, was evaluiert wird, wobei der Prozess des Organisierens vier bis fünf Schritte mit vielen Beteiligten umfasst.

**Geplante Änderungen der Evaluationspraxis** In Zukunft soll die Evaluation zielgerichteter werden. Zum einen sollen zu evaluierende Kurse gezielter ausgewählt werden, z.B. Vormittags- oder Abendkurse oder Kurse von bestimmten Lehrenden, zum anderen soll Evaluation zur Überprüfung von herangetragenen Problemen systematisiert werden.

**Unterstützungsbedarf** Externe Unterstützung wird aufgrund der Impulse von außen generell positiv gesehen, es gibt aber keinen konkreten Bedarf.

**Nutzen und Nutzung der Evaluation** Die Evaluation liefert der KVHS gute konkrete Hinweise zu räumlichen Bedingungen, zum Anmeldeverfahren und zu Werbeverfahren. Die Ergebnisse werden anhand einfacher, aussagekräftiger Grafiken in Gremien, z.B. im VHS-Beirat, in Außenstellensitzungen und auf Lehrendenkonferenzen vorgestellt. Die Mitarbeitenden genießen positive Rückmeldungen, z.B. bekommen die Mitarbeitenden der Verwaltung höhere Teilnehmendenzahlen mit. Die Reflexion über das Handeln und die Situation erhöht das Bewusstsein für die eigene Arbeit und dadurch deren Qualität. Kaum ein Aufwand sollte zu viel sein, um die positiven Effekte von Evaluation zu erzielen, aber sie muss durchführbar bleiben. Um den

Nutzen der Evaluation zu erhöhen, könnte man die Ergebnisse in weiteren Gremien vorstellen.

**Benötigte Kompetenzen für die Evaluation** Um Bildungsprozesse effizient zu evaluieren, sollte man Fragen in Fragebögen so stellen können, dass die Teilnehmenden die Fragen verstehen und passende Antworten geben können. Man benötigt außerdem eine Sensibilität dafür, welche Ergebnisse etwas bringen, und wie man Evaluation mit geringem Erkenntnisgewinn so strukturiert, dass man Ergebnisse erzielt, die weiterbringen und aus denen man Handlungen ableiten kann.

**Vorhandene Kompetenzen** Die benötigten Kompetenzen sind als Ergebnis eines längeren Prozesses jetzt in der KVHS vorhanden.<sup>74</sup>

**Verständnis von Evaluation** Evaluation dient laut I01 dazu, den Rahmen für Bildungsprozesse, der nicht nur durch reinen Lerninhalt geprägt ist, optimal zu gestalten. Leitung und andere pädagogische Mitarbeitende teilen laut I01 vermutlich dieses Verständnis.

#### 6.2.4 Kommerziell orientierte bzw. wirtschaftsnahe Organisation

**Organisation** Es handelt sich um eine große Organisation mit 41 hauptamtlich pädagogischen Mitarbeitenden. Im Jahr 2008 wurden 96 Veranstaltungen durchgeführt. Zielgruppe sind Mitarbeitende der Unternehmen in der Region, Arbeitssuchende, Schulabgänger für die höhere Berufsfachschule und Privatpersonen. Der Schwerpunkt der Bildungsangebote liegt auf der beruflichen Bildung: Unter anderem werden Aus- und Weiterbildung auf den Gebieten Technik und Technologie, Wirtschaft und Verwaltung, EDV und Multimedia angeboten. Die Organisation hat eine AZWV-Anerkennung<sup>75</sup> zur Durchführung von Maßnahmen der Arbeitsagentur und bietet als hauptsächliche Veranstaltungsformen überbetriebliche Erstausbildung und Maßnahmen der Arbeitsagentur in Ganztagsausbildung bzw. Seminaren sowie berufliche Weiterbildung in Abendseminaren an.

**Evaluationspraxis** Die Organisation berücksichtigt folgende Aspekte bei der Evaluation: allgemeine Fragen zur Person und zu Erwartungen, allgemeine Einschät-

---

74 I01 hat die Frage, was er in einem Seminar zu Evaluation lernen wollte, nicht konkret beantwortet. Da sich die Antwort nicht sinnvoll in das Case Summary integrieren ließ, enthält das Summary keinen Abschnitt zu diesem Thema.

75 AZWV steht für „Anerkennungs- und Zulassungsverordnung Weiterbildung“. Unter [www.azwv.de](http://www.azwv.de) finden sich weiterführende Informationen der Bundesagentur für Arbeit. Zum 01.04.2012 wurde AZWV umbenannt in AZAV „Akkreditierungs- und Zulassungsverordnung Arbeitsförderung“.

zung des Kurses bezüglich Raum, Inhalt, Gestaltung etc. sowie Einschätzung der Lehrenden. Als Evaluationsinstrumente kommen schriftliche Befragungen, strukturierte und unstrukturierte Interviews, maßnahmebegleitende Gesprächsrunden, angekündigte und unangekündigte Hospitationen zum Einsatz. Der Umfang der Evaluation ist abhängig von der Maßnahmedauer. Maßnahmen bis 40 Std. werden durch Feedbackrunden evaluiert. Bei Maßnahmen bis 80 Std. werden Gesprächsrunden und Hospitationen im Verlauf der Maßnahme und eine schriftliche Befragung am Ende verwendet. In Maßnahmen mit über 80 Std. finden modul- bzw. zeitraumbezogene Gesprächsrunden und eine schriftliche Befragung statt. Verantwortlich für die Evaluation sind die jeweiligen Leitenden der Maßnahme. Die Organisation verfolgt mit der Evaluation den Zweck, Maßnahmen, Lehrende und Umfeld zu bewerten, denn die Teilnehmendensicht ist gegebenenfalls nötig zur Korrektur des als richtig erachteten Wegs.

**Herausforderungen bei der Evaluation** Problematisch ist es, schriftliche Befragungsergebnisse richtig zu werten, denn bei langen Maßnahmen beeinflussen beispielsweise Inhalte und Vorkommnisse direkt vor dem Befragungszeitpunkt die Ergebnisse. Zudem kommt es darauf an, welches Klientel man befragt: Langzeitarbeitslose mit negativer Grundeinstellung geben schlechtere Bewertungen als freiwillige Teilnehmende einer beruflichen Weiterbildung. Einige wenige Teilnehmende sehen die Evaluation am Ende als überflüssig an, weil sie dann keinen Einfluss nehmen können. Nachdem die verwendeten Fragebögen inzwischen einen Überarbeitungsprozess durchlaufen haben, lassen sie sich gut von den Teilnehmenden beantworten und von der Organisation auswerten.

**Geplante Änderungen der Evaluationspraxis** In Zukunft soll die Aktualität der Fragebögen und der Instrumente überprüft und dabei kontrolliert werden, ob die häufig wechselnden Mitarbeitenden damit zurechtkommen.

**Unterstützungsbedarf** besteht keiner, denn es gibt ausreichend Unterstützung durch die externen Zertifizierer und die internen wissenschaftlichen Mitarbeitenden.

**Nutzen und Nutzung der Evaluation** Die Evaluationsergebnisse werden genutzt, um Dozenten auszuwechseln, die nicht mit dem Klientel zurechtkommen und um andere Medien auszuwählen und einzusetzen, die den Teilnehmenden näher sind. Über Evaluationsergebnisse werden die Zuständigen von Maßnahmen und vorgesetzte Leitungen, bei Problemen die höhere Ebene und gegebenenfalls die Geschäftsleitung informiert. Die Öffentlichkeit wird nicht informiert.

**Benötigte Kompetenzen für die Evaluation** Da zur Evaluation nicht nur die Erfassung, sondern auch die Auswertung gehört, muss man methodisch-didaktisch gebildet sein, sich mit dem Lehrprozess auskennen und wissenschaftlich arbeiten können. Selbstverständlich gehört auch „Rechnen können“ (I04:79) für statistische Auswertungen dazu.

**Vorhandene Kompetenzen** sind: „Na alle!“ (I04:83)

**Gewünschte Lerninhalte eines Seminars zur Evaluation** In einem Seminar wäre eine Übersicht über gute rechnergestützte Auswertungsmethoden hilfreich.

**Verständnis von Evaluation** Evaluation bedeutet für I04, Prozesse zu beobachten, Schlussfolgerungen abzuleiten und Konsequenzen zu ziehen. Besonders die Konsequenz ist wichtig, also dass etwas bei der Evaluation rauskommt. Die anderen Personen in der Organisation teilen laut I04 vermutlich diese Sicht.

### 6.3 Die Hauptproblembereiche

Mit den im vorigen Abschnitt vorgestellten Fallzusammenfassungen lag ein erster analytischer Zugang zum Datenmaterial vor, der den Fokus auf den Einzelfall richtet. Die Erstellung der Case Summaries diente unter anderem als Vorbereitung der sich anschließenden kategorienbasierten Auswertung, indem sie erste Auffälligkeiten sowie Hinweise auf sinnvolle Analysekategorien zu Tage brachte. Beginnend mit diesem Kapitel wird der Blickwinkel jetzt geweitet und es werden die Ergebnisse der kategorienbasierten Auswertung präsentiert, welche die erhobenen Informationen aller 24 Organisationen berücksichtigt.

Es ließen sich fallübergreifend sieben unterschiedliche Problembereiche identifizieren, mit denen sich nicht nur einzelne, sondern mehrere Organisationen bei der Evaluation ihrer Bildungsprozesse konfrontiert sehen. Es handelt sich also nicht bloß um Einzelaussagen, sondern um mehrfach anzutreffende Herausforderungen:

- Evaluationsinstrumente und -verfahren (von 16 Organisationen genannt)
- Datenerhebung (12 Organisationen)
- Akzeptanz und Belastung von Teilnehmenden, Lehrenden und Mitarbeitenden (12 Organisationen)
- Datenauswertung und -bewertung (11 Organisationen)
- Aufwand und Ressourcen (11 Organisation)

- Nutzen, Erkenntnisgewinn und Aussagekraft (10 Organisationen)
- Nutzung, Umsetzung und Ergebnisverwendung (7 Organisationen)

Die sieben Problembereiche werden in den folgenden sieben Unterkapiteln im Detail erläutert.

### 6.3.1 Evaluationsinstrumente und -verfahren

Die Analyse der Interviews bestätigt eindeutig das Analyseergebnis der Selbstrepor-te (vgl. Abschnitt 6.1): Das Hauptproblemfeld für die Organisationen stellen Herausforderungen in Bezug auf die verwendeten Evaluationsinstrumente dar. Ein gutes Drittel der Organisationen beschreibt in diesem Zusammenhang Probleme, die man dem Bereich des *Messens oder Operationalisierens* zuordnen kann. Auf einer generellen Ebene besteht beispielsweise die Herausforderung, wie man ein Instrument weiterentwickelt, „das dann wirklich auch das misst, was wir wissen wollen von den Teilnehmenden“ (I09:19). Konkret stehen die Organisationen zum einen vor der Frage wie sie das eigene, teils komplexe Bildungsverständnis in für Teilnehmende verständliche Fragen übersetzen können (I01:57) und wie sie bestimmte, für sie relevante pädagogische Bildungsziele (I02:118), Konstrukte wie die Ausbildungsfähigkeit (I05:21) oder die Gruppendynamik (I05:71) und Kenngrößen wie die Abbrecherquote operationalisieren bzw. überhaupt messen können:

„Oder Abbrecherquote zum Beispiel so, und da hatten wir jetzt die Auseinandersetzung, das war ganz hilfreich, sich dann noch einmal Gedanken darüber zu machen. Ist ja alles gut und schön, aber wie messen Sie das?“ (I18:66)

Mehrere Organisationsvertreter/innen antizipieren bezüglich des Messens auch gleich den zu erwartenden und den bisher erreichten Nutzen und weisen auf methodische Probleme des Messens und Operationalisierens hin. Sie bemängeln, nicht genau zu wissen, wie man fragen muss, um längere, aussagekräftigere Antworten auf offene Fragen zu bekommen (I20:31), wie man zu nutzbringenden Daten kommt (I18:49), und wie man an „die tatsächlichen Knackpunkte“ (I15:30) bei einer großen Masse an Daten kommt. Letztere Frage stellt sich insbesondere deshalb, weil das eingesetzte Instrument nicht als differenzierend genug erlebt wird:

„Also wir haben wunderbare Rückmeldungen in der Regel, bewegen uns im ‚gut‘ bis ‚sehr gut‘ Bereich ... . Ich habe das Gefühl irgendwie, das bestätigt uns ein-

fach erst mal alles, dass wir eine gute Arbeit machen, das reicht mir nicht, so.“  
(I15:30)<sup>76</sup>

Bezüglich des Messens wird auch als problematisch gesehen, wenn die standardisierten Ergebnisse zwar ein positives Ergebnis anzeigen, man als Lehrende/r aber dennoch mit einem unguten Gefühl aus der Bildungsveranstaltung geht, also ein Widerspruch zwischen subjektiver Einschätzung und den vorliegenden Evaluationsergebnissen besteht:

„Also, Sie können am Ende einer Veranstaltung eine erstaunlich gute standardisierte Evaluation kriegen und haben als verantwortlicher Leiter der Veranstaltung trotzdem das Gefühl, da ist irgendwas nicht so gut gelaufen. Ja in der Regel können Sie das auch benennen. Ja, also an der und der Stelle gab es einen Bruch oder ist eine Diskussion schlecht gelaufen, die weitere Auswirkungen auf die gesamte Veranstaltung hat oder ähnliche Dinge.“ (I23:20)

Als weiteres Problem des Messens und Operationalisierens ist die Erfassung von Langzeitwirkungen und des nachhaltigen Nutzens für die Teilnehmenden zu berichten, denn wie I02 (Abs. 112) betont: „Das ist sicherlich ein blinder Fleck.“ Für einige Organisationsvertreter/innen haftet ferner dem Messen und Operationalisieren etwas Unprofessionelles an. So seien die erhobenen Daten nicht empirisch valide, weil die Fragebögen keine Kontrollfragen enthielten (I23:32) oder der ganze Fragebogen sei nicht ausreichend an der Definition gelungenen Lernens orientiert, aus der man eigentlich Lernziele herausarbeiten und diese dann mit Indikatoren versehen müsste (I12:23).

Neben dem Messen und Operationalisieren fallen zahlreiche weitere Probleme bezüglich der Evaluationsinstrumente und -verfahren auf, die allerdings seltener genannt werden. Hierzu zählen die *Fragen im Fragebogen*, denn die Teilnehmenden würden einfache und verständliche Fragen benötigen (I04:97), da sie z.B. mit zu allgemeinen Fragen schwer zurechtkämen (I02:32). Insgesamt ist es also „schwierig so ein Bewertungssystem richtig, also zumindest die Fragen richtig zu formulieren“ (I20:39). Dies gilt insbesondere dann, wenn man versuchen möchte, „immer kreativ die Fragen so zu stellen, dass man, ja, vielleicht auch ein bisschen was bei den Dozenten rauskitzeln kann, ohne dass es zu auffällig ist, worauf man hinaus möchte“ (I10:66). Auch die *Anpassung des Inhalts von Fragebögen* ist als Problem zu nennen,

---

<sup>76</sup> Auslassungen in Literaturziten werden mit drei Auslassungspunkten „...“ gekennzeichnet, während die Kennzeichnung von Sprechpausen der Interviewpersonen in den Transkripten zusätzlich Klammern um die drei Punkte „(...)“ aufweist.

wenn beispielsweise die Organisationen vor der Herausforderung stehen, die gestellten Fragen im Fragebogen zu verfeinern, oder dass im Bogen Bereiche zusätzlich mit abgefragt werden sollen, zu denen bereits Kritikpunkte bekannt geworden sind (I12:33; I19:19). Hinzu kommt die *Anpassung eines Fragebogens an spezielle Teilnehmendengruppen* wie etwa Migrantinnen, die nur geringe Deutschkenntnisse mitbringen (I18:34) – ein Problem, das sich auch bei der Analyse der Selbstreporte herauskristallisiert hat.

Analog zur Analyse der Selbstreporte zeigt sich auch in den Interviews das Spannungsfeld zwischen *Vereinheitlichung und Spezialisierung von Instrumenten*. Einerseits muss der Fragebogen bei unterschiedlichen Kursen einsetzbar sein, da zu viele verschiedene Bögen zu aufwändig erscheinen (I04:97), andererseits passen die abgefragten Inhalte nicht gleich gut für alle Veranstaltungsformen und einzelne Kurse erhalten nur aufgrund der abgefragten Inhalte schlechtere Bewertungen:

„Also uns ist es immer wichtig, dass die Trainer auch mit praxisrelevanten Beispielen arbeiten. ... . Jetzt gibt es aber durchaus mal eine Veranstaltung, wo das gar nicht so im Vordergrund stehen kann. Ja, der Trainer muss zwangsläufig dann in diesem Bereich schlecht abschneiden, obwohl er eigentlich sehr gut war. Und dann müssten wir im Grunde genommen schon mehr auch differenzieren. Und das tun wir nicht. Und so gesehen, ist es nicht immer optimal gelöst. Aber ich denke im Schnitt doch gut, nach unserer Erfahrung.“ (I14:44)

Der letzte Satz des Zitats deutet an, dass dieses Spannungsfeld zwar als Problem für den Einzelfall konstatiert werden kann, durch die Aggregation von Daten in dieser Organisation jedoch zufriedenstellend ausgeglichen wird. Eine andere Organisation wählt hingegen den Weg der Spezialisierung der Instrumente und überlegt sich derzeit, „ob wir z.B. für den Sport- und Gesundheitsbereich einen eigenständigen Fragebogen entwickeln“ (I17:24).

Schließlich stellt sich bezüglich der Evaluationsverfahren auch ganz generell die Frage nach einem *passenden Instrument*, das bei der Qualitätsentwicklung weiterhilft und geeignet ist, die Entwicklung einer Organisation wirklich voranzutreiben (I09:19), oder mit dem sich Bedarfe der Teilnehmenden erschließen lassen (I10:19).

### 6.3.2 Datenerhebung

Während die Datenerhebung in den Selbstreporten nur als geringfügiges Problem zu Tage tritt, benennt immerhin die Hälfte aller Interviewpersonen Herausforderungen

in diesem Feld. Als Problem stellt sich zunächst der *geringe Rücklauf* heraus, insbesondere was Nachbefragungen anbelangt (I08:30), denn „wenn Leute sich bei uns abgemeldet haben, ist es teilweise schwierig die noch zu motivieren, dass sie sich nochmal mit uns beschäftigen“ (I13:42). Der geringe Rücklauf von Nachbefragungen führt auch zu Unzufriedenheit mit der Repräsentativität der erhobenen Daten (I19:19). Eng mit dem Rücklauf sind in einzelnen Organisationen Herausforderungen mit den *Lehrenden* bezüglich der Datenerhebung verbunden. So erscheint es in einer Organisation wichtig, dass man bei den Lehrenden wegen der Rückläufe „ein bisschen nachhakt“ (I17:22), da bei den Lehrenden unter Umständen auch Ängste vorhanden sind. Derartige Ängste können sogar die gesamte Datenerhebung beeinträchtigen, wenn z.B. Lehrende aus Angst, nicht weiterbeschäftigt zu werden, keine kritischen Rückmeldungen geben (I10:26).

Bei der Datenerhebung ist besonders ein Problemaspekt hervorzuheben, der indirekt auch noch einmal Fragen des Messens aufgreift. Und zwar die von den Organisationsvertreter/innen wahrgenommene Tatsache, dass unterschiedliche Teilnehmendengruppen auch unterschiedlich positiv antworten. Vereinfacht lassen sich die entsprechenden Interviewaussagen auf folgende (subjektive) „Gesetze“ reduzieren:

- Erzieher werten positiver als Lehrer (I02:70)
- freiwillig Teilnehmende werten positiver als verpflichtete Teilnehmende (I01:22); oder noch konkreter: freiwillige Teilnehmende einer berufsbegleitenden Weiterbildung werten positiver als Langzeitarbeitslose, die per se eine negative Grundhaltung haben (I04:39).

Diese *Verzerrungen der Daten* aufgrund von Grundeinstellung und Freiwilligkeit der Teilnahme beeinträchtigen auch die Auswertung und Bewertung der erhobenen Daten, weshalb man sich immer vor Augen führen muss, wie die Daten zustande gekommen sind, um Fehlinterpretationen zu vermeiden.

Einige wenige verweisen bezüglich der Datenerhebung auch auf den *Aufwand für die Teilnehmenden* bei der Datenerhebung, der entsteht, wenn beispielsweise Jugendliche jede Woche ein Lerntagebuch ausfüllen müssen (I05:31).

### **6.3.3 Akzeptanz und Belastung von Teilnehmenden, Lehrenden und Mitarbeitenden**

Eng verbunden mit dem letztgenannten Problem bei der Datenerhebung sind die Belastung der Teilnehmenden durch die Evaluation und die Akzeptanz der Evaluation



durch die Teilnehmenden. Zwischen der Akzeptanz und der Belastung besteht eine Wechselwirkung, die auch in weiteren Interviews zum Ausdruck kommt. Wenn die Akzeptanz gering ist und einige Teilnehmende am Ende der Maßnahme den Sinn der Evaluation nicht sehen, diese gar als überflüssig betrachten (I04:41), und man den Teilnehmenden es „erst plausibel machen [muss], dass sie jetzt gebeten sind bitte schön noch zur Auswertung etwas beizutragen“ (I09:27), dann kann die Evaluation auch zur Belastung ausarten. Wenn Teilnehmende ein und denselben Fragebogen in verschiedenen Veranstaltungen oder unterschiedliche Bögen in der gleichen Veranstaltung ausfüllen sollen, dann stellt dies eine Belastung der Teilnehmenden dar, die darauf mit Widerständen und wenig Akzeptanz reagieren können (I12:31; I17:26; I23:25).<sup>77</sup> Auch wenn Erhebungsbögen zu lang sind, können diese die Teilnehmenden ungebührend belasten, so dass umfangreiche Bögen gar nicht erst zum Einsatz kommen, auch wenn durch einen längeren Bogen die empirische Validität erhöht würde:

„Empirisch ist es nicht valide. ... Und wir haben uns dann auch Vorschläge machen lassen, wie man es valider machen könnte, aber haben das wieder verworfen, weil nämlich die Fangfragen ..., die man einbauen muss in einen solchen Fragebogen, um zu klären, ob das Antwortverhalten wirklich plausibel ist, die würden den Fragebogen natürlich weiter aufblähen von der Länge her. Und da würden wir echt ein echtes Problem sehen, also wir haben jetzt zwei Seiten, je nach Länge der Veranstaltung. Und der Vorschlag der dann da lag, das waren vier Seiten und das ist schon ein kleiner Unterschied im Hinblick auf die Zumutung gegenüber den Teilnehmenden.“ (I23:32)

Auch für die Gruppe der Lehrenden und Mitarbeitenden spielt die Belastung durch die Evaluation und die Akzeptanz der Evaluation eine Rolle – und auch hier lässt sich eine Wechselwirkung zwischen Belastung und Akzeptanz verzeichnen, denn wenn Lehrende verpflichtet sind Feedbackgespräche zu dokumentieren, steigt ihr Aufwand und es sinkt zugleich ihre Akzeptanz gegenüber der Evaluation (I18:42). In der Einrichtung von I17 zeigen sich zudem auch bei der Verwaltung Vorbehalte aufgrund der Arbeitsbelastung durch die Evaluation. In anderen Organisationen bestehen bei neuen Lehrenden oder auch bei langjährigen Mitarbeitenden Widerstände

---

<sup>77</sup> Hier ist direkt der Evaluationsstandard D1 „Angemessene Verfahren“ der DeGEval angesprochen, der da lautet: „Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen“ (DeGEval, 2008, S. 26).

gegen die Evaluation, welche als „von außen aufgesetzt“ erlebt wird, und schließlich „gibt es durchaus Lehrende, die in der Minderheit sind, aber man muss es trotzdem ernst nehmen, die sagen ‚Warum soll mich ein Teilnehmer bewerten?‘“ (I21:20). Nicht nur bei Teilnehmenden, auch unter Lehrenden kann sich Unmut breitmachen, wenn die Belastung durch wiederholte Befragungen steigt: „Ich bin doch gerade erst evaluiert worden oder ich bin doch schon im letzten Semester evaluiert worden, ist denn das wirklich nötig, dass das jetzt schon wieder geschieht?“ (I06:21).

Die Interviewaussagen legen nahe, zwischen einerseits der Akzeptanz der Evaluationsfrequenz und der damit verbundenen Belastung sowie andererseits der Akzeptanz der Evaluation als Ganzes zu unterscheiden. Während die Frequenz – wie oben ausgeführt – durchaus als eine Herausforderung für Teilnehmende und zum Teil auch Lehrende wahrgenommen wird, ist die Gesamtakzeptanz größtenteils gegeben. Wenn, dann gibt es vereinzelte Personen oder kleinere Personengruppen, die nicht von Evaluation „begeistert“ sind, was aber toleriert bzw. nicht als Problem gewertet wird:

„Und wir haben bestimmte Gruppen, die mögen das gar nicht, dass da ... die Qualität ... also gerade jetzt von eben Lehrenden, dass da die Qualität überprüft wird und rückgemeldet wird. Aber das sind eher kleinere Gruppen. So insgesamt gesehen.“ (I15:34)

Stellt man alle 24 Organisationen vergleichend nebeneinander, lassen sich drei Typen bezüglich der Akzeptanz erkennen.<sup>78</sup> Der erste Typ (11 Organisationen) benennt durchweg kein Problem bei der Akzeptanz unter Teilnehmenden, Mitarbeitenden und Lehrenden. Unter diesem Typus dominieren kleinere Organisationen sowie Organisationen, deren Bildungsangebot schwerpunktmäßig auf der beruflichen Bildung liegt, und Organisationen, die dem kommerziellen oder wirtschaftsnahen Bereich zuzuordnen sind. Der zweite Typ (8 Organisationen) sieht bei vereinzelten Gruppen, wie bestimmten Teilnehmenden, Lehrenden oder Mitarbeitenden, Probleme und der dritte Typ (4 Organisationen) macht tendenziell widersprüchliche Aussagen, etwa I02, der Evaluation als „Riesenherausforderung“ für einzelne Mitarbeitende beschreibt, wenn diese wissen, dass ihr Kurs evaluiert wird, und gleichzeitig bezüglich der Akzeptanz unter Mitarbeitenden kein Problem sieht.

---

<sup>78</sup> Ein Befragter (I01) ließ sich nicht zuordnen, da keine Informationen zu diesem Themenbereich vorliegen.

Obwohl fast die Hälfte der Organisationen dem Typus 1 zuzuordnen ist, soll nicht der Eindruck aufkommen, dass die Akzeptanz eine Selbstverständlichkeit darstellt. Zahlreiche Aussagen machen deutlich, wie diese erst im Laufe der Zeit erarbeitet werden musste. Unter den Lehrenden wurden „Ängste weggeräumt“, wurde für Verständnis geworben, es wurden die Vorteile herausgestellt und die Lehrenden wurden wiederholt zur Evaluation angehalten (u.a. I15:34; I17:26; I20:18). Man könnte auch sagen, es handelte sich um einen Sensibilisierungsprozess, in dem „man die Leute aufklären musste, wie wichtig das ist und wie wichtig das für die Organisation ist und wie wichtig das für sie selbst sein kann. Also jetzt für die Lehrenden“ (I24:49).

#### **6.3.4 Datenauswertung und -bewertung**

Bezüglich der Auswertung erhobener Daten lassen sich mehrere Problemaspekte erkennen, die größtenteils sehr einrichtungsspezifisch sind: So liegen Ergebnisse erst zeitverzögert vor, aus der verwendeten Datenbank lassen sich nicht alle gewünschten Daten ohne Weiteres und ohne großen Aufwand auslesen, die Verantwortlichkeiten sind nicht klar geregelt und es stellt sich als problematisch dar, wie man mit großen Datenmengen umgeht oder wie man die Ergebnisse unterschiedlicher Regionalstellen zusammenstellt (I06:21; I11:45-50; I13:60; I15:54-55; I24:50). Darüber hinaus stellt sich die Frage, wie man Daten so auswertet und sinnvoll aufbereitet, dass sie sowohl der Geschäftsleitung eine gute Übersicht in aggregierter Form liefert als auch den Lehrenden detaillierte Informationen über Erfolgsaspekte einzelner Bildungsangebote preisgeben (I18:51). Zudem führen – aufgrund einer selbst wahrgenommenen fehlenden Professionalisierung der Evaluation – unterschiedliche Berechnungsweisen bei der Datenauswertung zu unverlässlichen und nicht vergleichbaren Ergebnissen (I12:36-39), was in der Folge natürlich auch eine Bewertung der Ergebnisse verkompliziert.

Bezüglich der Bewertung von Daten stehen zahlreiche Einrichtungen vor grundlegenden Herausforderungen, die sich in folgenden Fragen ausdrücken lassen: Wie können wir die Ergebnisse verallgemeinern? Wie können wir die Ergebnisse interpretieren, wenn wir genau überprüfen müssen, ob die Teilnehmenden die Fragebogen-Fragen richtig verstanden haben? Wie bewerten wir Abbrecherquoten? Welchen Standards fühlen wir uns verpflichtet, welche Zielvorstellungen haben wir und welche können wir überhaupt erreichen? Bewerten wir es dennoch als Erfolg, wenn drei Viertel die (externe) Prüfung nicht bestehen? Wie nehmen wir eine Gesamtbe-

wertung vor, wenn wir viele unterschiedliche, im Grunde nicht vergleichbare Bildungsangebote haben? (I02:86; I16:18; I16:22; I18:48; I18:66). Eine weitere Frage besteht darin, inwieweit kurzfristige Ereignisse im Kursverlauf die Evaluationsergebnisse beeinflussen:

„Eine Herausforderung ist immer das, was man in einer schriftlichen Befragung erfragt, auch richtig zu werten. Denn es ist ganz oft so, dass die Teilnehmer bei einer Evaluation immer so die letzten Ereignisse im Gedächtnis hatten und dann können, wenn ein Kurs über 80 Stunden geht, die ersten 75 Stunden super gelaufen sein, und die letzten Stunden haben irgendein Vorkommnis gehabt und schon kann die Bewertung negativ sein oder in umgedrehter Richtung dann auch zu positiv. Also das abzuschätzen inwieweit da kurzfristige Ereignisse den Ausschlag gegeben haben, ist natürlich immer so die Herausforderung.“ (I04:37)

Während es sich bei der Datenauswertung eher um einrichtungsspezifische Probleme handelt, kristallisiert sich bei der Datenbewertung die Unsicherheit bei der Bewertbarkeit und Belastbarkeit der Daten als zentrales einrichtungsübergreifendes Problem heraus. Dieses übergreifende Problem betrifft letztendlich auch Gütekriterien wie die Validität der Datenerhebung. Sowohl die Herausforderungen bei der Datenauswertung als auch der Datenbewertung konzentrieren sich eher auf die mittelgroßen und großen Einrichtungen und werden nur von zwei Vertreter/innen der zehn kleinen Organisationen berichtet.

### **6.3.5 Aufwand und Ressourcen**

Evaluation ist anstrengend, aufwändig, schwer und stellt eine Zusatzbelastung dar, „weil es mit in das normale Tagesgeschäft rein muss“ (I05:95) – mit diesem überspitzt formulierten Satz, lassen sich die Aussagen zum Problembereich Aufwand und Ressourcen zusammenfassen. Konkret stellt die Bewältigung von „viel Masse“ (I15:30 u. 50), insbesondere bei den Fragebögen, einen erheblichen Arbeitsaufwand dar und es stellt sich die Frage, wie man einen guten Mittelweg findet, um einerseits zu Ergebnissen zu kommen und andererseits die beteiligten Personen zeitlich nicht zu stark zu belasten (I18:38). Nicht nur die Datenerhebung, sondern auch die zunehmend systematische Verwendung von Ergebnissen wird dabei als Arbeitsbelastung beschrieben:

„Also wenn ich jetzt aus jedem Seminar Rückmeldungen bekomme, auch schriftlich und auch kritische Rückmeldungen habe, muss ich mich ja darum kümmern. Und, was ja gut ist, aber (...) ja. Was aber möglicherweise dann eben Arbeits-

schwerpunkte verändert. Ich muss öfter mal mit einem Haus, mit einer Tagungsstätte reden oder mit Lehrenden reden und fragen, was ist denn da los gewesen. So, das war ja vorher [vor LQW] in dieser systematischen Form nicht erforderlich.“ (I15:42)

Ressourcen fehlen des Weiteren allgemein in Form von Geld oder bei der Auswertung, so dass man auf Qualifizierungsarbeiten von Studierenden angewiesen ist, bei denen für die Organisationen keine Kosten anfallen (I10:77; I24:50). Aufgrund fehlender Ressourcen und Kapazitäten lassen sich auch bestimmte Datenerhebungsmethoden, wie aufwändige Hospitationen oder leitfadengestützte Interviews, nicht umsetzen, wenngleich der Wunsch danach besteht (I13:48-52; I21:18).

Die Analyse der Selbstreporte offenbarte zahlreiche Probleme der Organisationen bezüglich des Aufwands von Evaluationen. Erstaunlicherweise wird demgegenüber in den Interviews der Aufwand im Verhältnis zum Nutzen von den meisten Interviewpersonen jedoch als angemessen beurteilt. Nur drei der Organisationsvertreter/innen empfinden den Aufwand als zu groß, von denen zwei davon ausgehen, dass dieses Missverhältnis rasch bzw. einfach umgekehrt werden kann. Die dritte Interviewperson stellt eine Verbindung zu den nicht ausreichend differenzierten Messinstrumenten her:

„Aber dieser ganze weitere Erfassungsaufwand und das was uns das bringt an Hinweisen auf Verbesserungspotenziale, das sind zu wenig eigentlich. Und da hätte ich gerne was, was uns da mehr bringt. Also mehr sagt, aussagekräftiger ist, differenzierter ist. Also im Grunde müssten wir irgendwie was machen, wenn wir feststellen, dass wir sowieso im ‚gut‘ und ‚sehr gut‘ Bereich sind ganz überwiegend, dann müssten wir uns jetzt irgendwas überlegen, wie wir den Bereich nochmal differenzierter herauskriegen.“ (I15:86)

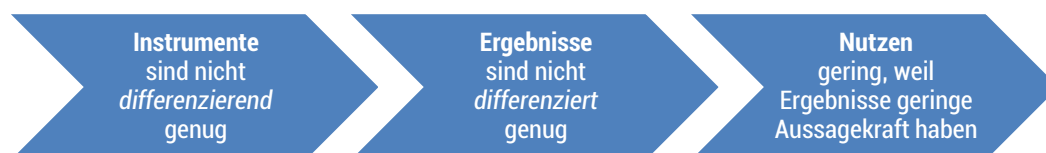
#### **6.3.6 Nutzen, Erkenntnisgewinn und Aussagekraft**

In diesem Problembereich sind Aussagen zusammengefasst, die noch nicht die konkrete Nutzung der Daten betreffen, wohl aber die Bewertung der Nützlichkeit der Ergebnisse ansprechen, wie etwa deren Aussagekraft und den durch sie erlangten Erkenntnisgewinn. Gerade letzterer stellt ein häufig betontes Problem dar. So liefert die Evaluation für I01 keine neuen Erkenntnisse aufgrund zu guter und zu wenig differenzierender Daten:

„Also wir haben ganz selten mal eher negative Äußerungen z.B. über den Kursleiter bekommen, von denen wir nicht ohnehin schon gewusst hätten, oder über andere Wege erfahren hätten. Also es sind keine neuen Erkenntnisse daraus gewonnen. Wenn wir was evaluiert haben, und haben dann bei einem Kursleiter z.B. eher negative Ergebnisse bekommen, dann haben wir meistens gedacht, ach ja, alles klar. Wir wissen auch schon, wer das ist.“ (I01:22)

Die Ergebnisse lassen sich nicht ausreichend für die Konzeptionierung von Bildungsmaßnahmen (I18:41) und für die Verbesserung der Lehr-Lern-Prozesse einsetzen und nutzen, weil unter anderem „die Daten einfach zu gut sind“ (I06:55). Die Ergebnisse sind also zu einem gewissen Grad nutzlos, was auch passieren kann, wenn eine Organisation viele Einzelveranstaltungen anbietet, deren Evaluationsergebnisse sich nicht auf anders konzeptionierte Folgekurse übertragen lassen (I07:21; I23:20). Die Evaluation bringt also keine Erkenntnis hervor, die „dann weiterhilft in der Qualitätsentwicklung“ (I09:19). Vielmehr ist das Gegenteil der Fall, denn „bestimmte Dinge oder ein Großteil der Aussagen ähnelt sich von Halbjahr zu Halbjahr. Das hat aber ein bisschen was damit zu tun, dass die Fragebögen selbst nicht genügend durchdacht sind“ (I12:41). Die im letzten Zitat angedeutete Kausalkette von Instrument, Ergebnis und Nutzen, die auch in anderen Interviews deutlich wird, ist in Abb. 43 veranschaulicht: Probleme bei den Instrumenten, wie etwa geringe Differenzierung bei der Rückmeldung durch die Teilnehmenden, ziehen wenig differenzierende Ergebnisse nach sich, was dazu führt, dass der Nutzen der erhobenen Daten, also etwa die Verwertbarkeit der Daten aufgrund geringer Aussagekraft, ebenfalls gering bleibt.

**Abb. 43: Kausalkette von Instrument, Ergebnis und Nutzen**



Auffällig ist, dass Probleme bezüglich des Nutzens, des Erkenntnisgewinns und der Aussagekraft anscheinend für Organisationen mit Schwerpunkt auf der beruflichen Bildung nur eine untergeordnete Rolle spielen, denn nur zwei Vertreter/innen der 13 Organisationen beruflicher Bildung berichten von diesbezüglichen Problemen.

### 6.3.7 Nutzung, Umsetzung und Ergebnisverwendung

Während im vorigen Problembereich der Nutzen der Evaluation und ihrer Ergebnisse beleuchtet wurde, stehen beim letzten großen Problembereich Herausforderungen im Vordergrund, welche bei der Umsetzung und Ergebnisverwendung, also bei der konkreten Nutzung der Evaluationsdaten auftreten. Konkret führe Routine dazu, dass man nicht so schnell reagiere oder Hinweise im Haus nur verzögert an die richtige Stelle weitergebe (I03:47), wichtig wäre aber, konsequent „dranzubleiben“ und sofort Konsequenzen zu ziehen (I11:36). Schwierig gestaltet sich auch der Erkenntnistransfer zu den Lehrenden:

„Ich glaube ..., dass wir aber auf der Ebene Dozenten/Lehrgangsleiter die Ergebnisse besser verwerten könnten. Weil wir arbeiten hier in dem Bereich viel mit Freiberuflern, ... weil das ja ganz unterschiedliche Fächer sind, die da unterrichtet werden. Und hier gelingt es nicht immer optimal, dass man die Ergebnisse dem Freiberufler so nahe bringt, dass er auch das ganze umsetzt und verbessert und den Teilnehmerbedürfnissen anpasst.“ (I08:54)

Als weitere Herausforderung stellt sich die Frage, wie man die Daten überhaupt nutzt, insbesondere für strategische Entscheidungen, beispielsweise mit welchen Partnern man weiter zusammenarbeitet (I15:30; I16:42).

In Abgrenzung zu konkreten Nutzungsproblemen liegen Probleme bei der Umsetzung und Ergebnisverwendung tendenziell einen Prozessschritt weiter. Hier stellt sich für einzelne Organisationen nicht die Frage, wie man die Ergebnisse nutzt, sondern wie man es schafft, die eruierten Konsequenzen auch tatsächlich umzusetzen, wer für diesen Prozess verantwortlich ist und sich auch verantwortlich fühlt und darum kümmert (I03:89; I09:21; I24:43).

### 6.3.8 Weitere Einzelprobleme

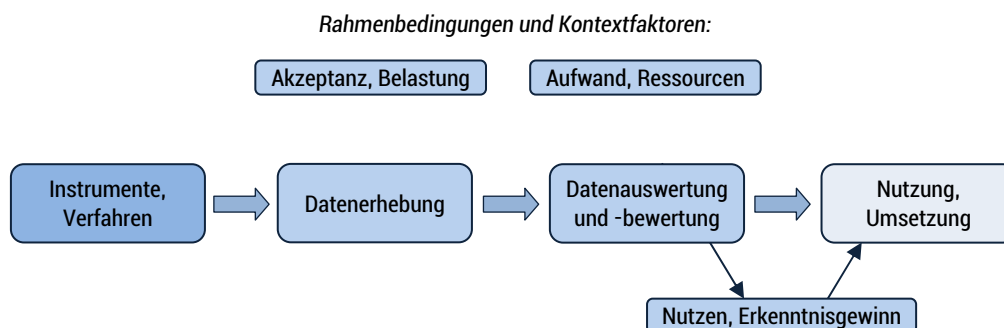
Neben den aufgeführten sieben großen Problembereichen existieren zahlreiche weitere Herausforderungen, die hier lediglich kurz dargestellt werden sollen, da sie jeweils nur von maximal vier Interviewpersonen genannt wurden. Zur *Organisation des Evaluationsprocedures* zählen die Probleme, regelmäßig an allen Standorten Evaluationen durchzuführen, die Evaluation kontinuierlich zu betreuen und unterschiedliche Evaluationen für einen Kurs inhaltlich zu koordinieren. Einige beklagen eine aufkommende *Routine und Evaluationsmüdigkeit* und andere sehen Probleme bezüglich der *vorhandenen Kompetenzen* bzw. weisen darauf hin, dass es zu einem

Problem fehlender Kompetenzen hätte kommen können, wenn nicht mit der Person des Befragten ein Sozialwissenschaftler im Hause gewesen wäre. Die *Ergebnisdarstellung und Informationsweitergabe* stellt für zwei Organisationen ein Problem dar, etwa im Sinne einer adäquaten Aufbereitung der Ergebnisse in Form einfacher und leicht verständlicher Grafiken für die Außendarstellung. Nicht unerwähnt bleiben sollen zum Schluss die ganz speziellen, nur von einzelnen Organisationen genannten Herausforderungen, die für die jeweiligen Einrichtungen zum Teil jedoch sehr bedeutsam sind. Hierzu zählen das Engagement der Mitarbeitenden und eine in Bezug auf Evaluation wenig motivierende Leitungskraft, unterschiedliche Ansichten zwischen Auftraggebern und Teilnehmenden bezüglich der Bildungsziele sowie der Bedarf an langfristigen Evaluationstätigkeiten sowie eines Gesamtkonzept für die Evaluation.

#### 6.4 Die sieben Hauptproblembereiche in der Gesamtschau

Abb. 44 zeigt noch einmal die in den Interviews identifizierten sieben Problembereiche, wobei die Darstellung der Probleme dem Ablauf eines Evaluationsprozesses folgt und zugleich deutlich macht, in welchem Bezug die Probleme zueinander stehen. So stehen die am häufigsten genannten methodischen Herausforderungen bezüglich der Instrumente und Verfahren zu Beginn des Evaluationsprozesses, während die Nutzung und Umsetzung dessen Abschluss bildet. Die Akzeptanz und die Belastung sowie der Aufwand und die Ressourcen stellen Rahmenbedingungen und Kontextfaktoren dieses Prozesses dar, während die Bewertung des Nutzens und des Erkenntnisgewinns die Nutzung und die Umsetzung beeinflusst.

Abb. 44: Die sieben Hauptproblembereiche verortet im Ablauf eines Evaluationsprozesses



Je dunkler ein Bereich dargestellt wurde, desto mehr Organisationen haben ihn benannt.



Wie bei der Darstellung der Herausforderungen im vorherigen Abschnitt an einigen Stellen deutlich wurde, lassen sich vereinzelte Zusammenhänge zu Faktoren wie dem Organisationstyp oder auch der Organisationsgröße identifizieren. Insgesamt betrachtet handelt es sich jedoch nur um sehr wenige und nur gering ausgeprägte Zusammenhänge. So lässt sich auch bezüglich unterschiedlicher Veranstaltungsformen, von denen man annehmen könnte, sie erzeugten unterschiedliche Probleme, nur folgender Zusammenhang feststellen: Unter den Organisationen, die mehrmonatige Weiterbildungen anbieten, berichten überdurchschnittlich viele von Herausforderungen bei der Datenerhebung. Ob die Einrichtungen eher fortlaufende Abendkurse oder halb- bis mehrtägige Blockveranstaltungen anbieten, spielt hingegen keine Rolle für die genannten Herausforderungen.

Bezüglich der Dauer des LQW-Einsatzes lassen sich ambivalente Auswirkungen feststellen. So ist einerseits auffällig, dass nur eine der Organisationen, welche das LQW-Qualitätsmodell bereits fünf Jahre oder länger einsetzen, Herausforderungen bezüglich der Nutzung und Umsetzung von Evaluationsergebnissen benennt. Allerdings ist andererseits festzustellen, dass die Vertreter/innen dieser Organisationen häufiger Herausforderungen bezüglich des Erkenntnisgewinns und der Nützlichkeit der Ergebnisse äußern. Ferner geht der längere LQW-Einsatz auch nicht mit weniger Herausforderungen einher, allerdings verschieben sich diese geringfügig: Alle Aussagen über Herausforderungen bezüglich der Erhebung von nachhaltigen Wirkungen stammen von Vertretern/innen aus Organisationen, die LQW fünf Jahre und länger einsetzen.

## **6.5 Entwicklungsperspektiven, Unterstützungsbedarf und Evaluationsverständnis**

### **6.5.1 Geplante Änderungen der Evaluationspraxis**

Zwei Drittel der Organisationen berichten von geplanten Änderungen der Evaluationspraxis in ihren Organisationen. An erster Stelle stehen wie auch bei den Problemfeldern die Instrumente. Zum einen sollen die *Instrumente verändert und angepasst* werden, das heißt, dass Einrichtungen ihre Fragebögen regelmäßig überprüfen, die Inhalte der Definition gelungenen Lernens stärker berücksichtigen und Fragebögen vereinheitlichen wollen. Bei einer weiteren Organisation steht eine Optimierung der Fragebögen an, um die Ergebnisse besser auswerten zu können (I04:52-45; I11:62; I14:54-58; I17:40; I20:38-43). Eine Interviewperson hebt in die-

sem Zusammenhang die Bedeutung aller Organisationsbereiche hervor, die zur gelungenen Bildung beitragen, und möchte diese auch im Fragebogen integrieren:

„Ja, noch andere Bereiche mit zu befragen, die wir bis jetzt nicht in unserem Fragebogen hatten. Wir haben ja z.B. eine externe Kinderbetreuung, externen Service usw. Die hatten wir gar nicht aufgenommen und dann ist es im Laufen doch aufgefallen, dass wir ja nun stärker trainieren, uns als Ganzes zu sehen und nicht nur alleine von den pädagogischen Mitarbeitern, die Seminare durchführen, zu denken, oder externen Referenten. Ja und das ist eins unserer nächsten Veränderungsvorhaben. Dass wir den Bogen praktisch ausweiten, die Evaluation auf alle die in diesem Haus in irgendeiner Weise irgendetwas für die Teilnehmenden oder mit ihnen tun.“ (I03:53)

Dieses Zitat illustriert nicht nur den konkreten Änderungswunsch bezüglich der Evaluationsinstrumente, sondern offenbart zugleich eine (in den Interviews selten anzutreffende) LQW-orientierte Sicht auf Bildung und ein breites Verständnis eines Bildungsprozesses, an dem sehr viele unterschiedliche Personen und Bereiche beteiligt sind. Die geplanten Änderungen an den Instrumenten beziehen sich nicht ausschließlich auf Fragebögen – es gibt darüber hinaus auch den Wunsch, bei Abbrecherbefragungen stärker auf qualitative Elemente zu setzen, weil man sich von diesen nützlichere Ergebnisse verspricht als von den bisherigen rein quantitativen Abbrecherquoten (I13:42-46).

Neben der geplanten Anpassung von Instrumenten spielt auch die *Neueinführung bzw. Wiedereinführung von Instrumenten* eine große Rolle. Dabei handelt es sich vor allem um die Entwicklung neuer Fragebögen für Lehrende, Teilnehmende oder auch für den Beratungsbereich sowie um die Einführung einer Nachbefragung (I02:58; I11:62; I15:58; I20:43):

„Und wir arbeiten an einem Instrument zur nachträglichen Nutzenerhebung. Also Nutzenbefragung, Telefoninterviews. Das ist zurzeit auch in Erarbeitung und wird im Laufe dieses Jahres auch erprobt.“ (I15:46)

Ein Organisationsvertreter betont, die früher erfolgreich eingesetzten Interviews wieder aufleben lassen zu wollen, denn „wir müssen die, also für uns selbst ... mit aufnehmen. Weil wir haben die richtig gefunden, als wir den LQW-Report gemacht haben. Und das steht einfach in den vier Jahren an“ (I21:50). Bei den langfristigen Kursen einer Organisation wird ein neues Verfahren erprobt, das stark in den Lehr-Lern-Prozess eingebunden ist und Jugendliche dabei unterstützt, ihre Ausbildungs-

fähigkeit selbst einzuschätzen und zu reflektieren, um möglichst schnell Defizite aufzuzeigen (I05:39).

Abseits von geplanten Änderungen bezüglich der Evaluationsinstrumente finden sich fast nur Einzelnennungen; lediglich Änderungen bezüglich des Evaluationsumfangs werden von immerhin drei Interviewpersonen angesprochen. Dieser soll in Zukunft reduziert und durch systematische Stichprobenauswahl verbessert werden (I01:24; I02:48; I17:40). Die weiteren Einzelnennungen zeugen von geplanten Änderungen organisationsspezifischer Natur. So soll etwa die Auswertung, aber auch die Ergebnisverwendung systematisiert werden, eine generelle Vereinheitlichung der Evaluationsverfahren stattfinden oder auch die Information über Evaluationsergebnisse ausgebaut werden, indem man in weiteren Gremien die Ergebnisse präsentiert (I01:44; I02:48; I18:52; I18:66). Eine Organisation plant, den „Evaluationsprozess als solchen“ (I06:29) zu evaluieren, also eine Meta-Evaluation durchzuführen, um Möglichkeiten des Umgangs mit Widerstand zu erkunden.

### **6.5.2 Unterstützungsbedarf bei der Evaluation**

Knapp die Hälfte der Organisationen äußert, keinen Unterstützungsbedarf bei der Evaluation zu haben, da unter anderem kompetente Personen im Haus waren oder aufgrund einer Fusion von zwei Organisationen sich die Mitarbeitenden gegenseitig unterstützen konnten (I03:55; I22:42). Die andere Hälfte greift vor allem einen Kerngedanken des LQW-Modells auf und betont auf die Frage nach dem Unterstützungsbedarf die Reflexion durch Außenstehende, denn Unterstützungsbedarf „ist hilfreich, weil man einfach die Außensicht braucht. Sonst rührt man im eigenen Saft“ (I13:48). Mithilfe des „fremden Blicks“ und der „externen Perspektive“ lassen sich die Betriebsblindheit und der eingefahrene Blick überwinden (I01:26; I17:46; I20:45; I21:57). Inhaltlich erhofft man sich Unterstützung bei der Entwicklung eines leistungsfähigen Instrumentariums und bei der methodischen Durchführung der Evaluation (I02:60-62; I12:59).

Obwohl grundsätzlich eine Unterstützung positiv bewertet wird, schrecken die Interviewpersonen teilweise vor den möglichen Kosten zurück und setzen vor allem auf kostenlose oder kostengünstige Unterstützung im Rahmen von Qualifikationsarbeiten an Hochschulen oder durch Kollegen/innen anderer Weiterbildungsträger (I10:77; I19:56). Drei Organisationen geht es nicht um inhaltlichen, qualitativen Support, sondern um personelle Aushilfe. Sie betonen den Unterstützungsbedarf bei

der „Bewältigung der Masse“ (I15:60), denn „intern ist das eine Frage der Ressourcen. Nicht eine Frage des Wissens“ (I21:57).

### 6.5.3 Gewünschte Lerninhalte eines Seminars zur Evaluation

Die Interviewpersonen wurden gefragt, was sie oder eine andere Person der Organisation in einem Seminar zum Thema Evaluation lernen wollen würde – vorausgesetzt, Zeit und Geld wären vorhanden. Bis auf eine Person, die gar keinen Bedarf an einem Seminar sieht, weil sie das Know-how hätte, um selbst ein solches durchzuführen, äußern alle anderen mehr oder weniger konkrete Lerninhalte. Erneut stehen an erster Stelle die *Instrumente, Verfahren und Methoden*, zu denen Äußerungen von knapp der Hälfte aller Organisationen vorliegen. Folgende Fragen spiegeln in kondensierter Form die gewünschten Lerninhalte wider: Welche anderen, evtl. neuen Verfahren gibt es und welche Grenzen und Einsatzmöglichkeiten sind mit ihnen verbunden? Wie lässt sich der Fragebogen verbessern und wie stellt man Fragen, um verwertbare Antworten zu bekommen?

Neben den Instrumenten, Verfahren und Methoden lassen sich noch weitere gewünschte Lerninhalte thematisch zusammenfassen, doch ist auffällig, dass es sich immer um Themen handelt, die nur für wenige Organisationen Relevanz besitzen. Die Organisationen sind folglich bezüglich der geäußerten Lernbedarfe als eher heterogen einzuschätzen. Immerhin fünf Interviewpersonen nennen als Lerninhalte für ein Seminar zum Thema Evaluation Aspekte des *Aufwands und der Effektivität*. Man möchte lernen, „wie man mit geringstmöglichem Aufwand ein möglichst hohes Ergebnis erzielt, also so das Perpetum Mobile dort für den Bereich erfindet“ (I07:65) oder wie man mit Zeit effektiver umgeht (I03:99). Weitere gewünschte Lerninhalte beziehen sich auf die *Auswertung*, die ebenfalls effizient und mit rechnergestützten Methoden ablaufen und zu verwertbaren Ergebnissen führen soll (I04:87; I05:85-87; I18:118), sowie auf die *abfragbaren Inhalte, verwendbaren Bewertungskriterien und Indikatoren*. Vereinzelt ist auch der Wunsch nach mehr Wissen und Können bezüglich *der Nutzung und des Umgangs mit Ergebnissen* und zur Gestaltung des Gesamtprozesses der Evaluation zu vernehmen.

Mehrere Organisationsvertreter/innen heben darüber hinaus in erster Linie – zum Teil neben konkreten Lerninhalten – die Möglichkeit eines Erfahrungsaustausches und eines Feedbacks von außen im Rahmen eines Seminars zum Thema Evaluation hervor. Der Wunsch dahinter liegt nahe: Man möchte die Betriebsblindheit beiseite-

legen und von Personen, „die sich da tagtäglich mit beschäftigen“ (I22:76), praxiserprobte Konzepte kennenlernen.

#### 6.5.4 Erhöhung des Nutzens

Gefragt danach, wie sich der Nutzen der Evaluation in der Organisation erhöhen ließe, zeigt sich ein vertrautes Bild. Wiederholt nehmen die Instrumente, Verfahren und Methoden für die Hälfte der Organisationen den zentralen Stellenwert ein, während andere Faktoren zur Nutzenerhöhung nur für eine kleine Gruppe an Organisationen bedeutsam sind. Bezüglich der Instrumente sind es im Wesentlichen die bereits bekannten Aspekte Fragebogenoptimierung, Neuentwicklung, Integration qualitativer Elemente, Nachbefragung sowie Spezialisierung und Vereinheitlichung, die genannt werden. Zu den anderen Faktoren zählen neben vielen Einzelnennungen noch die etwas häufiger genannte Weitergabe von Evaluationsergebnissen zur Information von Gremien und zu Werbezwecken und die zugleich als Herausforderung gesehene Nutzung der Ergebnisse (u.a. I06:59; I17:78).

#### 6.5.5 Verfolgte Evaluationszwecke

Die Organisationsvertreter/innen wurden gefragt, mit welchem Ziel in ihrer Organisation evaluiert wird, welche Zwecke also die Evaluation verfolgt. Dabei dominieren vor allem die *Verbesserung, Optimierung und Anpassung* der Bildungsprozesse (von 17 Organisationen genannt) sowie die *Erfolgskontrolle und Zielüberprüfung* (von 10 genannt). Im Gegensatz zu den Selbstreporten, bei denen sich die Erfolgskontrolle und Zielüberprüfung an erster Stelle der Evaluationszwecke gezeigt hat, überwiegt allerdings in den Interviews deutlich der Zweck der Verbesserung. Während zudem in den Selbstreporten als Evaluationszwecke *Bedarfsermittlung und Programmplanung, Wissens- und Erkenntnisgenerierung, Reflexion, Steuerung und Legitimation* genannt werden, spielen diese in den Interviews kaum eine oder gar keine Rolle. Umgekehrt fällt auf, dass in den Interviews als Ziel der Evaluation von immerhin drei Organisationsvertretern/innen der wirtschaftliche Erhalt der Organisation genannt wird: „Das letztendliche Ziel ist, dass das Unternehmen am Markt bleibt und nicht irgendwo den Bach runter geht“ (I07:39). Dieser Aspekt wird in den Selbstreporten gar nicht angesprochen.

Selbst wenn man die grundlegend verschiedene Form der Daten (langer Prozess der Verschriftlichung im Selbstreport vs. spontan mündlich im Interview) als mögliche Erklärung für Unterschiede heranzieht, ist es doch erstaunlich, dass keine Inter-

viewperson als Evaluationszweck die *Unterstützung des Lernens und der Didaktik* benennt, dass also Lernende durch die Evaluation beispielsweise die Möglichkeit zur Reflexion über eigene Lernerfolge erhalten. Laut mehrerer Interviewpersonen steht zwar als Ziel für deren Organisationen oben an, die Zufriedenheit der Lernenden zu erreichen, doch nur bei einer Interviewaussage wird eine Lernendenzentrierung im engeren Sinne deutlich: „Um bestmöglich die Teilnehmer, die wir bekommen, unabhängig davon, ob das Umschüler, ob das Auszubildende sind, ob das Teilnehmer an Lehrgängen sind, bestmöglich zu fördern. Um denen das Maximum zu geben“ (I20:47).

### 6.5.6 Evaluationsverständnis

Das Verständnis von Evaluation wird zwar von den Interviewpersonen mit teilweise sehr unterschiedlichen Worten umschrieben, doch lassen sich einige inhaltliche Gemeinsamkeiten herausarbeiten. Für den Großteil der Organisationsvertreter/innen stellt Evaluation eine Erfolgskontrolle dar und dient der Überprüfung der Zielerreichung, wie das folgende Zitat illustrieren mag:

„Evaluation kennzeichnet für mich, dass ich vorher, naja es geht ja um die Erfassung, im Grunde, des Erfolgs. Und Erreichung von Zielen. Also genau, erreiche ich die Ziele, die ich mir vorher gesteckt habe? Wie erfolgreich habe ich mit Blick darauf, also qualitativ weniger quantitativ, qualitativ, wie erfolgreich habe ich da gearbeitet?“ (I15:02)

Evaluation dient zudem für mehrere Organisationen der Bestandsaufnahme, Beobachtung und Erforschung und ebenfalls mehrere Interviewpersonen setzen Evaluation mit Bewertung gleich und sehen Evaluation beispielsweise als zielorientierten Bewertungsprozess (I17:94). Bei allen drei Aspekten (Erfolgskontrolle, Bestandsaufnahme, Bewertung) schließt sich, wie einige ausdrücklich betonen, auch das Schlussfolgern und das Ziehen von Konsequenzen an, denn es „muss schon irgendwas dabei rauskommen“ (I04:93).

In vielen Interviewaussagen kommt meist zusätzlich zu den genannten Aspekten eine sehr individuelle Sichtweise von Evaluation an, wie die folgende Aufzählung der zahlreichen Einzelnennungen zu verdeutlichen mag: Evaluation ist gleichzusetzen mit Auswertung; sie dient der Verbesserung, der Reflexion und der Wertermittlung einer Bildungsmaßnahme; sie eignet sich, um den Bildungsrahmen optimal zu gestalten; Evaluation ist Qualitätssicherung des Unterstützungsprozesses Bildung; evaluieren heißt, so eng wie möglich mit den betroffenen Personen zu kommunizieren

(u.a. I03:103; I10:125; I19:109). Interessanterweise korrespondiert ein vorherrschendes Evaluationsverständnis eines Organisationsvertreters bzw. einer Vertreterin nicht mit bestimmten genannten Herausforderungen.

Die deutliche Mehrheit der Organisationsvertreter/innen geht davon aus, dass die anderen Organisationsmitglieder ein ähnliches oder sogar gleiches Verständnis von Evaluation haben wie sie. Nur in etwa ein Drittel der Organisationen hätten laut Aussage der Vertreter/innen einzelne Personen oder Personengruppen andere Auffassungen von Evaluation und es würden vermutlich einige Mitarbeitende den Kontrollaspekt betonen, Lehrende eher auch den Arbeitsaufwand erwähnen und andere Organisationsmitglieder Evaluation sogar als lästig bezeichnen oder die Mitarbeitenden der Verwaltung mehr auf Zahlen und Fakten abheben (I02:139-146; I03:105; I11:122; I20:101).

## 6.6 Typologisierung der Organisationen hinsichtlich der Herausforderungen

Anhand der berichteten Herausforderungen lassen sich die Organisationen verschiedenen Gruppen zuordnen. Das bei der Zuordnung eingesetzte Verfahren orientiert sich am Ablaufschema der typenbildenden Inhaltsanalyse nach Kuckartz (2012) und wird im Folgenden in seinen Grundzügen erläutert.<sup>79</sup>

**Schritt 1 – Festlegung der Gruppierungsmerkmale** Bei der Gruppierung der Organisationen wurden alle codierten Textsegmente zu den Herausforderungen berücksichtigt, die durch die vorhergehende inhaltliche Strukturierung in sieben Problembereiche sowie sonstige Einzelnennungen kategorisiert vorlagen (vgl. Abschnitt 6.3):

- Instrumente, Verfahren
- Datenerhebung
- Akzeptanz, Belastung
- Auswertung, Bewertung
- Effizienz, Aufwand, Ressourcen

---

<sup>79</sup> Bei der Typenbildung und ihrer Beschreibung wurden soweit wie möglich die gebräuchlichen Gütekriterien, wie z.B. die Explizierung des Merkmalsraums, das Sparsamkeitsprinzip und die einmalige Fallzuordnung (vgl. Kuckartz, 2010b), sowie die methodologischen Regeln nach Kelle und Kluge (2010) beachtet.

- Nutzen, Erkenntnisgewinn
- Nutzung, Umsetzung
- Sonstige Herausforderungen

**Schritt 2 – Erstellung von Fallzusammenfassungen** Während der inhaltlich strukturierten Auswertung wurden zu fast allen codierten Textsegmenten verdichtende Kurzzusammenfassungen erstellt, die mit der Kommentarfunktion des Analyseprogramms MAXQDA festgehalten wurden. Für jede Interviewperson wurden die Kurzzusammenfassungen der jeweils codierten Herausforderungen zusammengestellt und um die Kurzzusammenfassungen zu den Themen „beabsichtigte Veränderungen“, „Unterstützungsbedarf“, „Nutzenerhöhung“ sowie „gewünschte Lerninhalte eines Seminars zur Evaluation“ ergänzt. Diese Zusatzinformationen dienten dazu, die Typenbildung durch weitere relevante Informationen zu unterstützen. Auf diese Weise ist für jede Interviewperson eine Fallzusammenfassung entstanden, die über die genannten Herausforderungen in komprimierter Form informiert. Abb. 45 gibt exemplarisch und ausschnittsweise die Zusammenfassung für I09 wieder. In der linken Spalte ist die Kategorie (inklusive ihrer Sortierungsnummer) zu sehen, zu der die betreffende Textstelle zugeordnet wurde, die rechte Spalte enthält die Kurzzusammenfassung der Textstelle. Die ersten beiden Zeilen stellen Herausforderungen dar, die vierte Zeile enthält ergänzende Informationen zum Bereich „Nutzenerhöhung“.

Abb. 45: Auszug aus einer Fallzusammenfassung für I09

Der Textstelle zugeordnete Kategorie	Kurzzusammenfassung der Textstelle
01 Instrumente, Verfahren\Operationalisieren und Messen	Wie misst man wirklich das, was man wissen will?
01 Instrumente, Verfahren\passendes Instrument	geeignetes Instrument finden oder weiterentwickeln, das weiterhilft bei der Qualitätsentwicklung
...	
09 Nutzenerhöhung\Instrumente, Verfahren, Methoden	Weiterentwicklung des Instrumentariums: Präzisierung von Fragestellungen im Auswertungsbogen
...	

**Schritt 3 – Erste Typologisierung der Fallzusammenfassung** Die erstellten 24 Fallzusammenfassungen wurden anschließend zu Gruppen geclustert, indem die Fälle hinsichtlich ihrer inhaltlichen Nähe und Distanz untersucht wurden und Fälle, die sich besonders ähnlich waren, zu Gruppen zusammengefasst wurden. Auf diese Weise entstanden zunächst fünf Gruppen.



**Schritt 4 – Erstellung der vorläufigen Typdefinitionen** Für jede der so entstandenen Gruppen wurde in zwei bis drei Sätzen festgehalten, was für sie typische Kennzeichen sind, beispielsweise, dass die Instrumente und Verfahren für alle Organisationen einer Gruppe eine herausstechende Herausforderung darstellen.

**Schritt 5 – Überprüfung der Gruppenzuordnung** Um die Güte der Typologisierung zu überprüfen und zu erhöhen, wurden die jeweiligen Fälle einer Gruppe noch einmal auf ihre interne Homogenität und externe Heterogenität untersucht. Bei Unsicherheiten bezüglich der Gruppenzuordnung wurden zusätzlich zu den Fallzusammenfassungen auch die Originaltextstellen, zum Teil sogar die Audioaufnahmen, herangezogen, um eine abschließende eindeutige Zuordnung zu ermöglichen. Als Ergebnis der Überprüfung wurde die Zahl der Gruppen auf vier reduziert, da sich in den Originaltextstellen zweier Organisationen, die aufgrund unsicherer Zuordnung zunächst eine eigene Gruppe bildeten, deutliche Ähnlichkeiten zu anderen Gruppen zeigten.

**Schritt 6 – Erstellung der endgültigen Beschreibung der Typen** Im Anschluss an die erfolgte Gruppeneinteilung wurden alle vier Typen und ihre Charakteristika detailliert beschrieben und durch Originalzitate aus den Interviews angereichert. Zu jedem Typ wurde eine zentrale Frage herausgearbeitet, die charakteristisch für diese Gruppe an Organisationen ist und deren Fokus auf die Herausforderungen bei der Evaluation widerspiegelt. Alle Typen haben zudem ein kennzeichnendes Label erhalten. Sofern sich auffällige Zusammenhänge zu Faktoren wie der Organisationsgröße, dem Organisationstyp, der Funktion der Interviewperson etc. finden ließen, wurden diese – aufgrund der teilweise kleinen Fallzahl in den Gruppen mit Bedacht – in die Beschreibungen integriert.

Als Ergebnis der Gruppierung liegen vier Typen vor, die sich gemäß des Gruppierungsziels hinsichtlich der fokussierten Herausforderungen bei der Evaluation unterscheiden.<sup>80</sup>

---

80 Es handelt sich um eine natürliche, nicht um eine künstliche Typologie, die induktiv aus dem Datenmaterial gebildet wurde. Die vier Typen sind polythetisch, da die einzelnen Fälle innerhalb der Typen nicht gänzlich identisch hinsichtlich der Herausforderungen bei der Evaluation sind, sondern die Fälle eines Typs untereinander Variationen aufweisen und sich nur besonders ähnlich sind (vgl. Kuckartz, 2012). Es gibt gewissermaßen ein jeweiliges Typenzentrum, von dem die einzelnen Fälle des Typs näher und weiter entfernt liegen.

### 6.6.1 Typ I: Instrumente und Verfahren im Fokus

*Zentrale Frage: Wie gestalten wir die Instrumente und Verfahren? (n=13, Interviews: 1, 2, 4, 5, 9, 10, 12, 14, 15, 18, 19, 20, 23)*

Für die Organisationen dieser Gruppe stellen Evaluationsinstrumente und -verfahren eine wesentliche Herausforderung dar: Wie gestaltet man empirisch valide Instrumente? Wie formuliert man Fragen im Fragebogen, die leicht verständlich sind oder die das Erkenntnisinteresse nicht sofort verraten? Wie misst man Abbrecherquoten, die eigenen Bildungsziele oder die wirklichen „Knackpunkte“? Spezialisiert oder verallgemeinert man die Instrumente? Derartige Fragen würden die Organisationen, die die Instrumente und Verfahren bei den Herausforderungen fokussieren, auch gerne in einem Seminar zur Evaluation behandeln wollen, etwa welche standardisierten Verfahren es gibt und wie sich diese einsetzen lassen.

Die Organisationen dieser Gruppe beschreiben nicht nur aktuelle Herausforderungen bezüglich der Verfahren und Instrumente, sondern nennen häufig auch Entwicklungsperspektiven in diesem Bereich. So sollen etwa Fragebögen für Lehrende und Teilnehmende entwickelt und eine Nachbefragung installiert werden, aber auch die Anpassung von Fragebögen durch einen stärkeren Rückbezug auf die Definition gelungenen Lernens steht auf dem Entwicklungsprogramm.

Neben den Instrumenten und Verfahren werden häufig auch zusätzlich Probleme bei der Datenauswertung und -bewertung sowie – als Konsequenz der in Abb. 44 demonstrierten Kausalkette – Herausforderungen bezüglich des Nutzens und der Nutzung genannt. Im Unterschied zu Typ II, der primär die Ergebnisse fokussiert, betrachtet Typ I das Problem der Nutzung von der Warte der Instrumente aus und stellt zu diesen eine direkte Verbindung her:

„So und wir haben das jetzt so, wir wählen zwei bis vier Ziele aus, die bei der Veranstaltung im Vordergrund gestanden haben. Und die sollen dann bewertet werden von ‚Stimmt völlig‘ bis ‚Stimmt gar nicht‘ und da stellen wir fest, dass die Leute das nicht verstehen oder, also da können wir mit den Aussagen nicht so viel anfangen.“ (I02:36)

„Also z.B. Unzufriedenheit mit den Räumen, mit den Unterrichtsräumen, ist ein Problem, was immer wieder auftaucht, und das weiß man vorher schon (...), weil wir wissen einfach an welchen Unterrichtsorten die Räume nicht in Ordnung sind. Also bestimmte Dinge oder ein Großteil der Aussagen ähnelt sich von

Halbjahr zu Halbjahr. Das hat aber ein bisschen was damit zu tun, dass die Fragebögen selbst nicht genügend durchdacht sind.“ (I02:41)

„Also ich denke mal, wenn das ein allgemeines Seminar wäre, Evaluation von Bildungsprozessen, dass uns sehr helfen würde, wenn uns einer sagen würde, wie stelle ich am besten die Fragen, um eine verwertbare Antwort zu bekommen.“ (I20:95)

Man könnte vermuten, dass besonders zu Beginn der Einführung von LQW die Instrumente und Verfahren im Vordergrund bei den Herausforderungen stehen, doch lässt sich diese Hypothese nicht bestätigen, denn diesbezüglich lassen sich keine Auffälligkeiten im Datenmaterial finden.

### **6.6.2 Typ II: Ergebnisse im Fokus**

*Zentrale Frage: Wie nutzen wir die Evaluationsergebnisse effizient? (n=5, Interviews: 3, 7, 11, 16, 24)*

Im Gegensatz zum ersten Typ benennt der zweite Typ nicht oder nur am Rande die Instrumente und Verfahren als Problem, sondern die Organisationen „steigen“ im Evaluationsprozess erst später mit ihrer Problemsicht ein, nämlich bei der Datenbewertung und -auswertung. Zentrale Herausforderungen sehen die Organisationen dieser Gruppe jedoch vor allem bei dem Nutzen, den die Evaluation erbringt, sowie bei der Nutzung und Umsetzung von Evaluationsergebnissen. Die Einrichtungen stehen beispielsweise vor den Problemen, dass standardisierte Daten über viele unterschiedliche und wenig vergleichbare Bildungsangebote wenig bewertbar und aussagekräftig sind, dass Ergebnisse nicht konsequent und kontinuierlich weitergegeben und genutzt werden und dass die Nutzung der Ergebnisse eher spärlich und informell geschieht:

„Also ich finde der Nutzen wäre eigentlich groß, man müsste ihn nur noch ein bisschen anders aufbereiten und ein bisschen ernster nehmen und so.“ (I24:90)

„Die Schwierigkeit ist vielleicht darin, dann wirklich immer wieder konsequent zu gucken nach den Ergebnissen, wie sieht es denn aus. Dranzubleiben auch, um dann sofort Konsequenzen zu ziehen.“ (I11:36)

Die Fokussierung auf die Ergebnisse zeigt sich bei dieser Gruppe auch bei der Frage, welche Inhalte sie im Rahmen eines Seminars zur Evaluation lernen wollen würden:

„Also ich denke wichtig zu lernen wäre tatsächlich notwendige und richtige Schlussfolgerungen zu ziehen aus Ergebnissen. Ergebnisse zusammenzuführen und dann so einen Soll-Ist-Vergleich zu machen um dann eingreifen zu können in Prozesse. Mutig eingreifen zu können.“ (I11:114)

### 6.6.3 Typ III: Aufwand im Fokus

*Zentrale Frage: Wie bewältigen wir (auch noch) die Evaluation? (n=2, Interviews: 6, 17)*

Die Organisationen dieser Gruppe, beides Volkshochschulen, sehen als Herausforderung vorwiegend die Belastung und den Arbeitsaufwand, die mit Evaluation einhergehen. Hinzu treten auch das Problem der Akzeptanz der Evaluation und die Belastung durch die Evaluation. Lehrende monieren, sie seien schon letztes Semester evaluiert worden, und Teilnehmende sagen, sie müssten am Ende des Seminars schnell zum Bus und hätten keine Zeit mehr, den Evaluationsbogen auszufüllen. Auch bei den Mitarbeitenden sind Akzeptanzprobleme zu verzeichnen:

„Bei den Beschäftigten hier bei uns ist das so, dass da wo ausgewertet wird, erst quasi Grobauswertung gemacht wird, das schon als Arbeitsbelastung gesehen wird. Weil da die normalen Arbeitsprozesse immer wieder neu unterbrochen werden.“ (I17:26)

Angesichts derartiger negativer Assoziationen zu Evaluation, möchte eine der beiden Organisationen sogar eine Meta-Evaluation durchführen, um Möglichkeiten zum Umgang mit dem Widerstand zu erkunden:

„Also in dem Bereich der Kundenkommunikation ... haben wir uns drauf festgelegt, dass wir den Evaluationsprozess als solchen nochmal evaluieren. Also wenn Sie wollen, eine Meta-Evaluation machen. Das ist auch nötig, also wenn man schon über die ersten Widerstände redet, ist die Frage, wie geht man damit um? Macht man eine Motivation, macht man noch mal deutlich wozu das ganz dient, auch den Teilnehmern gegenüber, oder aber weicht man ab von teilstandardisierten Fragebögen und geht auf andere Methoden über?“ (I06:29)

#### 6.6.4 Typ IV: Konzeption im Fokus

*Zentrale Frage: Wie organisieren und konzeptionieren wir die Evaluation als Gesamtprozess? (n=4, Interviews: 8, 13, 21, 22)*

Den Organisationen, die im Typ IV zusammengefasst wurden, ist allen gemeinsam, dass sie weniger einen Blick auf einzelne Detailprobleme werfen, sondern vielmehr auf die Evaluation als Gesamtes schauen und dabei organisatorische Aspekte, grundlegende Dilemmata oder andere fundamentale Probleme (jenseits von Aufwand von Akzeptanz) benennen. So zählen beispielsweise zu den zentralen Herausforderungen dieser Gruppe, dass die Auftraggebenden von Bildungsmaßnahmen ganz andere – teilweise konträre – Bildungsziele verfolgen als die Teilnehmenden, dass langfristige Evaluationen notwendig seien, dass die Regelmäßigkeit der Durchführung an allen Standorten gewährleistet sein muss und dass ein Gesamtevaluationskonzept vonnöten ist:

„Und daher sind die Evaluationsanforderungen, die Herausforderungen der Evaluation, bestehen die nicht alleine darin einen Kurs zu evaluieren. Also weil der Unterschied jetzt zur Volkshochschule, die Volkshochschule macht einen Kurs und hat ein Publikum, was sie auch definieren kann, aber zu dem sie ansonsten nicht in einer vertraglichen oder Arbeitsbeziehung steht. Und das heißt, wir müssen bei der Evaluation immer ein Gesamtevaluationskonzept haben.“  
(I21:16)

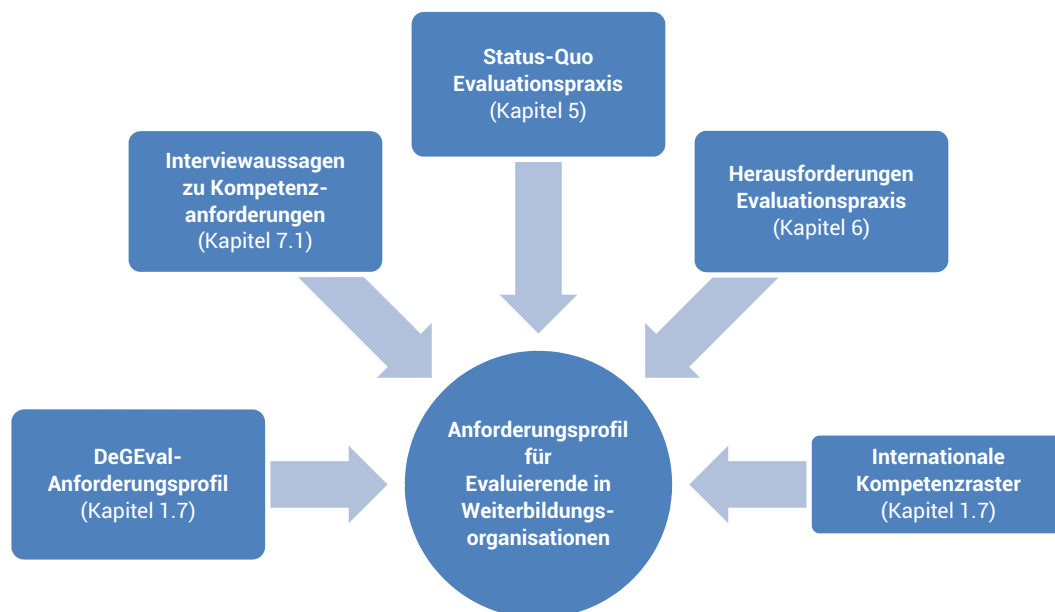
Für die Organisationen dieser Gruppe spielt das Gesamtprocedere der Evaluation die Hauptrolle und man könnte annehmen, dass vorwiegend Interviews mit Leitungspersonen geführt wurden, deren Aufgabe es in der Regel ist, den Gesamtprozess der Evaluation und seine Konzeption im Blick zu behalten. Dies ist aber nicht der Fall, denn nur zwei der vier Konzeptionsfokussierten sind Leitungspersonen.



## 7 Schlussfolgerungen: Ein Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen

Im siebten Kapitel wird ein Anforderungsprofil an Evaluierende in Weiterbildungsorganisationen vorgestellt, welches die bisherigen theoretischen Ausführungen und die Studienergebnisse zur Evaluation von Bildungsprozessen in Weiterbildungsorganisationen zusammenführt. Als Ausgangspunkt für die Entwicklung des Anforderungsprofils dienten die „Empfehlungen für die Aus- und Weiterbildung in der Evaluation – Anforderungsprofile an Evaluatorinnen und Evaluatoren“, die von der DeGEval herausgegeben wurden (Caspari et al., 2008) und mit denen ein seit mehreren Jahren etablierter Kompetenzkatalog vorliegt (vgl. Abschnitt 1.7.2). Dieser Kompetenzkatalog wurde unter Berücksichtigung verschiedener Analyseergebnisse adaptiert und konkretisiert. Abb. 46 zeigt, aus welchen Quellen Aspekte und Informationen in das Anforderungsprofil mit eingeflossen sind.

Abb. 46: Quellen für das Anforderungsprofil für Evaluierende von Weiterbildungsprozessen



In einem ersten Adaptionsschritt wurden die (bisher noch nicht vorgestellten) Antworten der interviewten Vertreter/innen von Weiterbildungsorganisationen auf die Frage, welche Kompetenzen man benötige, um Bildungsprozesse effizient zu evaluieren, mit dem Anforderungsprofil der DeGEval abgeglichen (vgl. Abschnitt 7.1). In einem weiteren Adaptionsschritt wurden die einzelnen Kompetenzfelder des DeGEval-Profiles in Bezug zu den in Kapitel 5 und 6 vorgestellten empirischen Ergebnissen

zum Status quo und zu den Herausforderungen der Evaluationspraxis gesetzt. Bei der Adaption wurden die bereits existierenden Anforderungen zum einen konkretisiert und zum anderen wurden Aspekte, die wenig Entsprechung in den empirischen Ergebnissen finden, durch Anforderungen ersetzt, die im bisherigen Profil unberücksichtigt sind, für die Durchführung von Evaluationen in Weiterbildungsorganisationen jedoch sehr relevant erscheinen. Darüber hinaus flossen zum Teil auch Anregungen aus internationalen Kompetenzrastern (vgl. Abschnitt 1.7.3) in das Profil mit ein. Als Resultat liegt ein Anforderungsprofil vor, das sich an den Bedarfen und Problemstellungen der Evaluationspraxis in Weiterbildungsorganisationen orientiert. Da für die Analyse Weiterbildungsorganisationen ausgewählt wurden, welche dem LQW-Modell folgen, wurden bei der Entwicklung des Anforderungsprofils auch die Spezifika dieses Modells berücksichtigt. Das Profil ist jedoch bewusst so formuliert, dass es problemlos auch auf Weiterbildungseinrichtungen übertragen werden kann, die nicht das LQW-Modell einsetzen.

Um ein angemessenes Kompetenzraster formulieren zu können, war es wichtig sich vor Augen zu führen, für welchen Personenkreis dies geschieht und in welchem Kontext dieser Evaluation betreibt, denn unterschiedliche Evaluationskontexte und unterschiedliche Evaluierende implizieren unterschiedliche Anforderungen an die Evaluierenden (vgl. Rädiker & Niklesz, 2012). Wie die Auswertung der Evaluationspraxis in Kapitel 5 gezeigt hat, lassen sich die von den Organisationen durchgeführten Evaluationen der Bildungsprozesse fast ausschließlich als interne bzw. Selbstevaluation klassifizieren. Da die Kompetenzanforderungen der DeGEval nicht speziell für den Bereich der Selbstevaluation formuliert sind, galt es diesen Aspekt bei der Adaption des Anforderungsprofils zu berücksichtigen. Ferner war zu beachten, dass es sich nicht (nur) um *die Selbstevaluation der Lehr-Lern-Prozesse durch Lehrende* handelt, sondern viel häufiger um *die Selbstevaluation der Bildungsprozesse durch pädagogisch, planende Mitarbeitende*, die zum Teil auch lehrend tätig sind. Das Anforderungsprofil geht zudem davon aus, dass die Evaluation der Bildungsprozesse nicht als Haupttätigkeit ausgeführt wird, sondern als ein Aufgabenfeld unter vielen in die pädagogische Arbeit integriert ist.

Wie in Abschnitt 1.7 dargelegt, liegt dieser Arbeit ein dreigeteilter Kompetenzbegriff zugrunde. So werden im Folgenden unter Kompetenzen das Wissen (im Sinne von Kenntnissen, *knowledge*), die Fertigkeiten (im Sinne von praktischen Fähigkeiten, *skills*) und die Haltungen (im Sinne von Einstellungen und Motiven, *dispositions*) ge-



fasst, das bzw. die für die Evaluation in Weiterbildungsorganisationen nützlich und zumeist notwendig sind.

Das entwickelte Anforderungsprofil erfüllt mehrere Zwecke. So wie die Kompetenzanforderungen der DeGEval als Grundlage für Weiterbildungsprogramme in der Evaluation dienen sollen (vgl. Caspari et al., 2008), so bietet auch das hier formulierte Anforderungsprofil eine Grundlage für die Entwicklung eines Curriculums oder Fortbildungskonzepts für eine spezielle Zielgruppe und es lassen sich beispielsweise Lernziele für ein konkretes Bildungsangebot ableiten. Dabei können die in Kapitel 6 herausgearbeiteten Problem-Typen sowohl dabei helfen, die Curricula von Bildungsangeboten speziell auf bestimmte Gruppen zuzuspitzen, als auch diese gezielt zu bewerben. Zum Beispiel sind für Organisationstypus I „Instrumente und Verfahren im Fokus“ mit Sicherheit Angebote zu Methodenfragen von großem Interesse, während Workshops, die den Evaluationsprozess behandeln, größeren Anklang beim vierten Problemtyp „Konzeption im Fokus“ finden dürften. Allerdings sollte beachtet werden, dass für die jeweiligen Organisationstypen nicht zwangsläufig die Fortbildungsinhalte optimal sind, die dem ersten Anschein nach sinnvoll erscheinen – nämlich dann, wenn eine Organisation bildlich gesprochen einen „blinden Fleck“ bezüglich ihrer Evaluationspraxis hat. Denn es ist ja nicht auszuschließen, dass in einer Organisation der starke Fokus bei der Evaluation auf Methodenfragen zum Teil eine Folge von mangelnder Evaluationskonzeption und -planung ist, etwa indem der konzentrierte Blick auf die Probleme der Fragebögen das viel grundlegendere Problem der unbefriedigenden Stakeholderbeteiligung überdeckt.<sup>81</sup> Über die Konzeptionierung von Bildungsangeboten hinaus kann das Anforderungsprofil als Instrument zur Personalentwicklung in Organisationen genutzt werden, indem über vorhandene und benötigte Evaluationskompetenzen reflektiert wird.<sup>82</sup> So heben Zech et al. (2010, S. 186-212) die Bedeutung von Schlüsselkompetenzen im Rahmen des Personalmanagements hervor und beschreiben ein Kompetenzmodell, das vier Kompetenzdimensionen unterscheidet, die als Grundlage der Handlungskompetenz

---

81 So betonen Beywl et al. (2007, S. 8) die große Bedeutung der Planung einer Evaluation für eine Reduzierung des Aufwandes bei der späteren Datenauswertung, -bewertung und Ergebnisverwendung.

82 Dass Kompetenzen des Weiterbildungspersonals eine aktuell bedeutsame Rolle spielen, zeigt auch das Thema der Netzwerkkonferenz der Lerner- und Kundenorientierten Qualitätsentwicklung im Jahre 2012. Unter dem Titel „Kompetenzen begreifen – Mit Kunst die Zukunft gestalten!“ widmet sich die Tagung drei zentralen Fragen: (1) Welche Kompetenzen benötigen Bildungsorganisationen, um ihre Zukunftsfähigkeit zu sichern? (2) Wie können Bildungsorganisationen die für sie notwendigen Kompetenzen ermitteln? (3) Was können Bildungsorganisationen von künstlerischem Handeln lernen?

betrachtet werden: die Persönlichkeitskompetenz, die Sozialkompetenz, die Methodenkompetenz und die Fachkompetenz. Zech et al. legen zudem eine Anleitung vor, wie Weiterbildungsorganisationen ein organisationsspezifisches Kompetenzmodell entwickeln können. In Abb. 47 ist ein solches organisationsspezifisches Modell exemplarisch dargestellt, welches in der Organisation ArtSet®<sup>83</sup> zum Einsatz kommt und in allen vier Bereichen konkrete Einzelkompetenzen auflistet.

**Abb. 47: Auszug aus dem organisationsspezifischen Kompetenzmodell von ArtSet®**  
(Quelle: ArtSet, 2012, S. 55)

Das ArtSet®-Kompetenzmodell	
<b>Fachliche Kompetenz</b> <ul style="list-style-type: none"> <li>• theoretische und methodische Forschungskompetenzen</li> <li>• pädagogische und didaktisch-methodische Kompetenzen</li> <li>• Beratungskompetenzen</li> <li>• Begutachungskompetenzen</li> <li>• Managementkompetenzen</li> <li>• etc.</li> </ul>	<b>Persönliche Kompetenzen</b> <ul style="list-style-type: none"> <li>• Überzeugendes Auftreten</li> <li>• Leistungswille</li> <li>• Eigeninitiative</li> <li>• Analyse- und Problemlösefähigkeit</li> <li>• Ergebnisorientierung</li> <li>• Flexibilität</li> <li>• Selbstorganisationsfähigkeit</li> <li>• Belastbarkeit</li> <li>• Kreativität</li> <li>• unternehmerisches Denken und Handeln</li> <li>• Humor</li> <li>• etc.</li> </ul>
<b>Methodische Kompetenzen</b> <ul style="list-style-type: none"> <li>• Projektmanagement</li> <li>• Moderationsmethoden</li> <li>• Didaktische Methoden</li> <li>• Zeitmanagement</li> <li>• Visualisierungsmethoden</li> <li>• Präsentation</li> <li>• Prozesssteuerung</li> <li>• Evaluation</li> <li>• Beherrschung elektronischer Medien</li> <li>• etc.</li> </ul>	<b>Soziale Kompetenzen</b> <ul style="list-style-type: none"> <li>• Kommunikationsfähigkeit</li> <li>• Kontaktfähigkeit</li> <li>• Führungsfähigkeit</li> <li>• Kritik- und Konfliktfähigkeit</li> <li>• Beobachtungs- und Feedbackfähigkeit</li> <li>• Fähigkeit, Andere zu begeistern</li> <li>• Wertschätzung</li> <li>• Durchsetzungsvermögen</li> <li>• Fähigkeit, auf Andere einzugehen</li> <li>• etc.</li> </ul>

Das in diesem siebten Kapitel vorgestellte Anforderungsprofil buchstabiert im Detail die „Evaluationskompetenzen“ aus, die in Abb. 47 im unteren linken Kasten unter den „Methodischen Kompetenzen“ verortet werden.

<sup>83</sup> ArtSet® ist die Organisation, welche das LQW-Modell vor gut zehn Jahren entwickelt hat.

## 7.1 Von Organisationsvertretern/innen berichtete Kompetenzanforderungen

Die im Rahmen der empirischen Studie interviewten 24 Organisationsvertreter/innen wurden gefragt, welche Kompetenzen man benötige, um Bildungsprozesse effizient zu evaluieren. Die Antworten auf diese Frage lassen sich den einzelnen Kompetenzfeldern der DeGEval zuordnen. Abb. 48 zeigt, wie häufig die Kompetenzfelder und ihre jeweiligen Dimensionen von wie vielen Vertretern/innen genannt worden sind. Das fünfte Kompetenzfeld zur „Praxis der Evaluation“ wurde dabei nicht berücksichtigt, weil es primär die Durchführung von Praktika fokussiert und als querliegend zu den ersten vier Bereichen verstanden wird (vgl. Caspari et al., 2008, S. 25).

**Abb. 48: Abgleich der DeGEval-Kompetenzfelder mit den Aussagen der Organisationsvertreter/innen**

Kompetenzbereich	von x Organisationsvertretern/innen (n=24) genannt
<b>1. EVALUATIONSTHEORIE UND -GESCHICHTE</b>	Summe = 1
a) Grundlagen	1
b) Evaluationsgeschichte	--
c) Evaluationsansätze	--
d) Evaluationsstandards	--
<b>2. METHODENKOMPETENZEN</b>	Summe = 12*
a) Grundzüge der empirischen Sozialforschung und Untersuchungsdesign	3
b) Datenerhebung	10
c) Statistische Kenntnisse	3
d) Datenverarbeitung, -auswertung, -interpretation	1
e) Projektorganisation	--
<b>3. ORGANISATIONS- UND FELDKENNTNISSE</b>	Summe = 13*
a) Organisationswissen	--
b) Rechts- und Verwaltungswissen	--
c) Spezifische Feldkenntnisse	5
<i>Pädagogische Kompetenzen**</i>	11
<b>4. SOZIAL- UND SELBSTKOMPETENZEN</b>	Summe = 8
a) Soziale Kompetenz	--
b) Kommunikative Kompetenz	1
c) Kooperationskompetenz	--
d) Selbstmanagementkompetenzen	--
e) Lern- und Problemlösekompetenz	7

\* Die Häufigkeit ist kleiner als die Summe der darunterliegenden Zeilen, weil einige Vertreter/innen mehrere Dimensionen genannt haben.

\*\* Diese von den Vertreter/innen genannte Dimension ist kein Bestandteil des DeGEval-Anforderungsprofils, lässt sich jedoch von ihrem Inhalt her eindeutig dem Kompetenzfeld „Organisations- und Feldkenntnisse“ zuordnen.

Auf den ersten Blick fällt die starke Ungleichverteilung der Kompetenzfelder auf: Mit nur einer Nennung spielt das Feld „Evaluationstheorie und -geschichte“ nahezu gar keine Rolle, was unter anderem sicherlich darauf zurückzuführen ist, dass Praxisvertreter/innen befragt wurden. Dennoch ist diese geringe Anzahl überraschend, da diesem Feld nicht nur theoretische (vermeintlich praxisferne) Aspekte zugeordnet sind, sondern auch einfache Grundlagen wie etwa Evaluationsdefinitionen.

Die „Methodenkompetenzen“ werden hingegen von fast jedem zweiten genannt, was angesichts der in Kapitel 5 häufig und facettenreich berichteten Probleme bezüglich der Evaluationsinstrumente und -verfahren weniger verwundert. Die Organisationsvertreter/innen verweisen hierbei vor allem auf zahlreiche notwendige Fähigkeiten bezüglich der Datenerhebung, die im Anforderungsprofil der DeGEval unter anderem auch die Entwicklung von Erhebungsinstrumenten umfassen.

Die hohe Bedeutsamkeit der „Organisations- und Feldkenntnisse“ ist insbesondere auf eine Dimension zurückzuführen, die zwar kein Bestandteil des DeGEval-Anforderungsprofils ist, sich jedoch eindeutig diesem Kompetenzfeld zuordnen lässt und mit „Pädagogische Kompetenzen“ überschrieben werden kann. Diese pädagogischen Kompetenzen stellen die Konkretisierung des Anforderungsprofils für die Evaluation von Weiterbildungsprozessen dar – was erklärt, warum sie überhaupt genannt werden. Erstaunlich ist, dass fast die Hälfte aller Organisationsvertreter/innen diese pädagogischen Kompetenzen als Voraussetzung für die effiziente Evaluation von Bildungsprozessen sieht und dabei so zahlreiche Facetten anspricht: Man sollte eine Vorstellung des Lehr-Lern-Prozesses sowie von Bildungsprozessen haben; man sollte mit Methodik- und Didaktik-Kenntnissen aufwarten können, um unter anderem verstehen zu können, was Teilnehmende in ihrem Feedback überhaupt mitteilen wollen; man sollte ganz generell mit Teilnehmenden umgehen können und man sollte sozialpädagogische Kenntnisse besitzen, um die Entwicklung von Jugendlichen im Rahmen einer Evaluation fachgerecht beurteilen zu können.

Im Feld der „Sozial- und Selbstkompetenzen“ lassen sich die meisten von den Vertreter/innen genannten Kompetenzen der Dimension „Lernen“ zuordnen. Während die Lernkompetenz im Anforderungsprofil der DeGEval jedoch vorwiegend mit Blick auf die Rolle von externen bzw. Fremdevaluierenden formuliert ist, äußern die interviewten Personen auch Kompetenzen, die sich speziell aus ihrer Rolle als interne bzw. Selbst-Evaluierende ergeben. Als zentral sind hier vor allem die Änderungsbereitschaft und die Kritikfähigkeit zu nennen, also beispielsweise die Rückmeldungen der Teilnehmenden einerseits hören zu wollen und andererseits aushalten zu kön-

nen. Zudem spielt auch die Reflektionsfähigkeit, im Sinne einer objektiven Sichtweise und Distanzierungsfähigkeit, eine Rolle. Die Äußerungen der Organisationsvertreter/innen zu den Lernkompetenzen lassen sich schließlich noch dahingehend unterscheiden, ob sich das Lernen einerseits auf die Individualebene oder andererseits auf die Organisationsebene im Sinne sich weiterentwickelnder Evaluationsprozesse und -abläufe bezieht.

Die Organisationsvertreter/innen haben zahlreiche weitere Kompetenzanforderungen genannt, die sich keinem der vier Kompetenzfelder eindeutig zuordnen lassen, aufgrund ihrer häufigeren Nennung hier aber kurz beschrieben werden: Zunächst handelt es sich um eine *Führungskompetenz*, die für eine effiziente Evaluation von Bildungsprozessen als notwendig betrachtet wird. Man sollte demnach in der Lage sein, Vertrauen bei Teilnehmenden und Lehrenden zu erzeugen, um überhaupt das Einverständnis zu bekommen, eine Evaluation durchzuführen. Führungskompetenz sei auch notwendig, wenn es um den Umgang mit Ergebnissen ginge, etwa wie man Evaluationsergebnisse dazu nutzt, bei den Beschäftigten Entwicklungen anzuregen – statt sie bloßzustellen. Auch die *Kompetenz zur Ergebnisumsetzung* wird mehrfach genannt und umfasst die Fähigkeiten, Schlussfolgerungen ziehen zu können, Ergebnisse in die Organisationsentwicklung einzubinden, Wichtiges von Unwichtigem zu unterscheiden und zu wissen, in welcher Form und in welchen Bereichen Ergebnisse überhaupt vorliegen sollten und hilfreich sind. Darüber hinaus wird die Kompetenz benannt, sich in die Lernenden hineinversetzen zu können – eine Fähigkeit, die grundlegend für die Anwendung des LQW-Modells ist und im Begriff der „Lernerorientierung“ zum Ausdruck kommt.

## **7.2 Das Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen im Detail**

Der Abgleich der DeGEval-Kompetenzfelder und ihrer Dimensionen mit den Aussagen der Organisationsvertreter/innen zu den benötigten Kompetenzen zeigt deutliche gegenseitige Lücken auf. Das im Folgenden vorgestellte Anforderungsprofil erweitert deshalb das DeGEval-Profil um mehrere Dimensionen, reduziert das Profil um weniger bis gar nicht relevante Dimensionen und nimmt einige Konkretisierungen vor, indem Kompetenzfelder und Dimensionen umbenannt werden, wobei wie eingangs geschildert insbesondere die in den Kapiteln 5 und 6 vorgestellten Ergebnisse zum Status quo und zu den Herausforderungen berücksichtigt werden. Das fertige Anforderungsprofil umfasst die folgenden fünf Kompetenzbereiche:

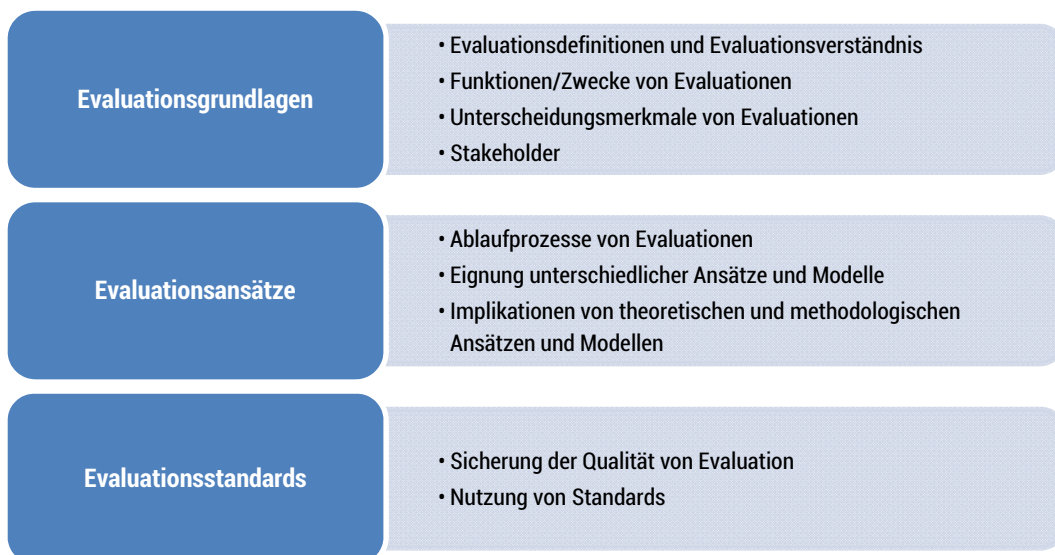
1. Basiswissen
2. Methodenkompetenzen
3. Organisationskenntnisse und pädagogische Kompetenzen
4. Lern- und Entwicklungskompetenzen
5. Nutzungskompetenzen

Im Folgenden werden die fünf Kompetenzbereiche und ihre jeweiligen Dimensionen im Detail vorgestellt.

### 7.2.1 Basiswissen

Der erste Kompetenzbereich bildet das Fundament für die folgenden Bereiche und umfasst das Basiswissen, das für die Durchführung von Evaluationen von Weiterbildungsprozessen grundlegend notwendig erscheint. Im Unterschied zum DeGEval-Kompetenzfeld „Theorie und Geschichte der Evaluation“ konzentriert sich der hier formulierte Bereich vorwiegend auf direkt handlungsrelevantes Wissen und hält bewusst historische Aspekte außen vor. Das Basiswissen gliedert sich auf in die Dimensionen a) Evaluationsgrundlagen, b) Evaluationsansätze sowie c) Evaluationsstandards.

Abb. 49: Dimensionen des Kompetenzbereichs „Basiswissen“



**Evaluationsgrundlagen** Wichtig erscheint zunächst, dass die Evaluierenden ihr Verständnis von Evaluation klar definiert haben und kommunizieren können und sich über die Konsequenzen unterschiedlicher Definitionen bewusst sind, etwa welchen Stellenwert das Bewerten in ihrer Auffassung von Evaluation einnimmt. Auch

die Abgrenzung und das Zusammenspiel mit verwandten Konzepten wie Monitoring und Bildungscontrolling sollte zum Grundlagenwissen der Evaluierenden gehören. Eng mit dem Evaluationsverständnis verknüpft sind die unterschiedlichen Funktionen bzw. Zwecke, die mit der Evaluation verfolgt werden, und die man unterscheiden können sollte, etwa diejenigen, die in den Selbstreporten am häufigsten genannt wurden: Zielüberprüfung, Optimierung, Erkenntnisgenerierung sowie Lehr- und Lernunterstützung. Dabei erscheint es zusätzlich hilfreich, den häufig mit Evaluation in Verbindung gebrachten Kontrollaspekt vor Augen zu haben. Um Evaluationen konzipieren zu können, ist es weiterhin wichtig, die Drehschrauben von Evaluationen zu kennen, also verschiedene Evaluationsmerkmale wie formativ und summativ sowie Fremd- und Selbstevaluation unterscheiden zu können – um nur ein paar wenige zu nennen. Das Wissen über potenzielle Stakeholder, insbesondere die Lernenden, Lehrenden, Planenden und Leitenden, und ihre jeweiligen Wünsche und Hoffnungen in Bezug auf die Evaluation ermöglicht den Evaluierenden eine nutzungsorientierte Ausrichtung der Evaluation. Aus diesem Wissen ergibt sich zudem eine Fähigkeit, die man mit „Stakeholder-Beteiligungs-Kompetenz“ überschreiben könnte.

**Evaluationsansätze** Als Grundlage für die Konzeption und Organisation von Evaluationen ist es hilfreich, einen allgemeinen Ablaufprozess von Evaluationen vor Augen zu haben, der die einzelnen Phasen von Evaluationsprojekten beschreibt und somit einen Orientierungsrahmen bietet. Unterschiedliche, für den Bereich Weiterbildung besonders relevante Evaluationsansätze und -modelle in ihrer Eignung für unterschiedliche Anwendungszwecke einschätzen zu können, gehört ebenfalls zum Basiswissen. Dies setzt natürlich einen Überblick über verschiedene Ansätze und Modelle voraus, wenngleich hierauf nicht der Schwerpunkt liegen sollte, denn weniger die Evaluationsansätze selbst als deren jeweilige Implikationen für die Praxis sind besonders relevant: Beispielsweise bedarf es nicht einer mehrtätigen Fortbildung in der Transformativen Evaluation (vgl. Mertens, 2008), wohl aber erscheint es sinnvoll, in einem Kurs Grundgedanken des Ansatzes darzulegen<sup>84</sup>, etwa dass Evaluation auch zum Empowerment der Teilnehmenden beitragen kann und Evaluationsmethoden nicht nur aus dem standardisierten Kanon empirischer Sozialforschung stammen müssen (vgl. Weber, 2012). Auch der Grundgedanke einer konsequenten Nutzenorientierung, wie ihn Patton in seinem Werk „Utilization-focused

---

84 Hier wird absichtlich nicht die Formulierung „vermitteln“ gewählt, um der Sichtweise Rechnung zu tragen, dass es eine Differenz zwischen Aneignung und Vermittlung in der (Erwachsenen-)Bildung gibt.

evaluation“ 2008 in vierter Auflage darlegt, könnte ein gewinnbringendes Wissen für die Organisationen darstellen. Ebenso stellt die Fähigkeit, den Programmbaum von Beywl et al. (vgl. Abschnitt 2.6.3) auf eine konkrete Bildungsmaßnahme in Grundzügen anzuwenden, sicherlich eine hilfreiche Kompetenz dar. Wenngleich der Ansatz der responsiven Evaluation nicht primär als Selbstevaluationsansatz formuliert wurde, so können doch wesentliche Elemente und insbesondere zugrundeliegende Denk- und Sichtweisen großen Gewinn für die Evaluation in LQW-Organisationen erbringen. Hierzu zählen insbesondere (vgl. Stake, 1973, 2004; Holla, 2002, S. 142-162):

- die Berücksichtigung der Eigenschaften des Lehrprozesses,
- die Einbeziehung von Expertenwissen sowie qualitativen Daten,
- die kritische Haltung gegenüber reinen Outcome-Messungen und die Betonung des intrinsischen Werts von Bildungsaktivitäten,
- die dezidierte Berücksichtigung unterschiedlicher Werteperspektiven der Beteiligten und die Akzeptanz multipler Realitäten sowie „uncertainty“,
- die stärkere Fokussierung auf Programmaktivitäten als auf „program intents“,
- die Einplanung von Zeit für informelle Ergebnisberichte, für die Eruierung von Teilnehmenden-Bedürfnissen, für die Sammlung von Bewertungen;

also alles in allem die Betonung und Berücksichtigung reflexiver Momente bei der Evaluation, man könnte auch sagen „Evaluation durch Reflexivität“.

**Evaluationsstandards** Evaluationsstandards dienen der Sicherung der Qualität von Evaluationen; ein Wissen über die Standardbereiche und ihre Inhalte hilft, die eigene Evaluationspraxis an diesen Standards auszurichten. Besonders hilfreich erscheint für die Evaluierenden, die im Fokus dieses Anforderungsprofil stehen, das Wissen über die Standards für Selbstevaluation der DeGEval (2004) sowie über die vereinzelt vorliegenden Veranschaulichungen und Erläuterungen der Evaluationsstandards (und ihrer Fallstricke). Solche Veranschaulichungen und Erläuterungen finden sich beispielsweise als Konkretisierung der Standards für die Schule bei Beywl (2010a) oder im „Handbuch der Evaluationsstandards“ des Joint Committee on Standards for Educational Evaluation, übersetzt von Beywl und Widmer (2006).



### 7.2.2 Methodenkompetenzen

Der zweite Kompetenzbereich umfasst das weite Feld der Methoden und bildet – aufgrund der Ergebnisse der empirischen Analysen – den Schwerpunkt des Anforderungsprofils. Wie zu Beginn des Kapitels angedeutet, lassen sich die von den Organisationen durchgeführten Evaluationen von Bildungsprozessen als Bestandteil professioneller Praxis charakterisieren und grenzen sich dadurch eindeutig von Evaluationsforschung als Typ sozialwissenschaftlicher Forschung ab (vgl. zu dieser Abgrenzung Lüders, 2006, S. 49). Diese Verortung der Evaluationen impliziert eine Orientierung der Methodenkompetenzen an praktischen Umsetzungsfragen und weniger an wissenschaftstheoretischen Fragen. Die Methodenkompetenzen teilen sich, wie bei der DeGEval, in zahlreiche Dimensionen auf: a) Planung und Durchführung empirischer Untersuchungen, b) Datenerhebung, c) Datenverarbeitung, -auswertung, -interpretation sowie d) Projektorganisation.

**Abb. 50: Dimensionen des Kompetenzbereichs „Methodenkompetenzen“**



**Planung und Durchführung empirischer Untersuchungen** Zentralen Stellenwert bei der Planung und Umsetzung von Evaluationen nehmen die Fragestellungen ein, die mit einer Evaluation beantwortet werden sollen. Daher erscheint es wichtig, Fragestellungen entwickeln und sie in ihrer Qualität und Reichweite beurteilen zu können. Dabei ist für Evaluierende insbesondere wichtig im Blick zu behalten, nur die Aspekte zu evaluieren, die im Einflussbereich der Weiterbildungsorganisation liegen und von ihr verändert werden können. Um den Aufwand von Evaluationen steuern zu können, ist es darüber hinaus wichtig, den Umfang und den Rhythmus der Evaluationen bewusst zu lenken. Hierbei sind insbesondere Kenntnisse über sinnvolle Auswahlverfahren zur Beschränkung des Erhebungsaufwandes notwendig. Mögliche Zeitpunkte der Evaluation (vor, während, am Ende, nach Abschluss eines Bildungsangebots) und die jeweils damit verbundenen Vorteile, Nachteile und Probleme sollten den Evaluierenden bewusst sein. Die Steuerung von Umfang, Rhythmus und Erhebungszeitpunkten sollte auf Basis der Abschätzung des Erkenntnisgewinns erfolgen, denn wiederholte Evaluationen, die keine neuen Erkenntnisse liefern, haben nur einen geringen bis gar keinen Nutzen. Wissen über Gesetze zum Datenschutz, insbesondere bezüglich Anonymität und Vertraulichkeit, ist notwendig, um die Rechte der Teilnehmenden zu wahren.

**Datenerhebung**<sup>85</sup> Es gibt zahlreiche Datenerhebungsformen und -instrumente, vom klassischen und häufig eingesetzten Fragebogen über Interviews und Tests bis hin zu Verfahren, die direkt in den Lehr-Lern-Prozess eingebunden sind. Vor dem Hintergrund dieser Vielfalt scheint es notwendig, diese zu kennen, aber vor allem auch die jeweiligen Möglichkeiten, Grenzen und Einsatzfelder einschätzen zu können. Wichtige Aspekte für die Einschätzung sind der Aufwand bezüglich Datenerhebung und -auswertung, die Belastung der Teilnehmenden und Lehrenden sowie die Antizipation von Ertrag, Erkenntnisgewinn sowie Nutzen und Nutzung, um nicht nur die oftmals beklagten „Datenfriedhöfe“ zu produzieren. Gleich, welches Datenerhebungsinstrument man wählt, wird man geeignete Indikatoren auswählen und gegebenenfalls entwickeln können müssen, die das abbilden, was man messen und erheben möchte. Da Fragebögen nicht nur sehr beliebt sind, sondern auch eine viel erprobte und erforschte Datenerhebungsmethode darstellen, erscheint es notwendig, mit fundierten Anwendungskenntnissen über ihre Entwicklung aufwarten zu kön-

---

<sup>85</sup> Das Verständnis der Datenerhebung in diesem Anforderungsprofil orientiert sich an dem Verständnis der DeGEval und ist umfassender als bei der zuvor durchgeführten Analyse (vgl. Kapitel 5 und 6). So umfasst es unter anderem auch die Instrumente und Verfahren.

nen. Dieses Know-how umfasst alle wichtigen Grundlagen der Fragebogengestaltung, wie die Fragenformulierung (das sogenannte „Wording“) und die Verwendung von Antwortskalen, und ist in zahlreichen Publikationen beschrieben, wie etwa bei Porst (2009) oder mit speziellen Hinweisen für die Evaluation bei Reischmann (2006) und Rädiker (2009). Gleichsam ist es notwendig, Datenerhebungsformen jenseits der Fragebögen entwickeln und einsetzen zu können, beispielsweise geeignete Fragen für persönliche Interviews formulieren, aber auch Lernstandskontrollen sowie Beobachtungsbögen konstruieren zu können. Die Verknüpfung von verschiedenen Erhebungsformen sowie von qualitativen und quantitativen Daten verspricht einen großen Gewinn, ist jedoch gleichsam voraussetzungsreich in ihrer konkreten Umsetzung: „Wie geht man mit Widersprüchen in den Daten um?“ oder „Wie lassen sich die Daten überhaupt aufeinander beziehen?“ sind zwei exemplarische Fragen für dieses komplexe Feld der Datentriangulation und Methodenkombination, die Evaluierende in Weiterbildungsorganisationen beantworten können sollten. Unabhängig, ob man Daten bzw. Methoden miteinander verschränkt oder nicht, sollten Evaluierende die Kontextabhängigkeit von Daten berücksichtigen und reflektieren können, denn jede Erhebung von Daten geschieht zu einem bestimmten Zeitpunkt (z.B. direkt nach einer aktivierenden Übung im Seminar) und in einer bestimmten Situation (z.B. durch die erstmalig im Seminar erscheinende Leiterin der Bildungseinrichtung). Evaluierende in Weiterbildungsorganisationen sollten darüber hinaus in der Lage sein, einfache Wirkungsanalysen sowie Abbrecherbefragungen und Nachbefragungen von Teilnehmenden konzipieren und durchführen zu können. Da die Lehrenden eine Mittlerfunktion zwischen Organisation und Teilnehmenden einnehmen, kommt ihnen meist auch eine zentrale Rolle bei der Datenerhebung zu. Die Evaluierenden müssen daher imstande sein, die Lehrenden in den Prozess der Datenerhebung mit einzubeziehen und ihnen Verantwortung zu übertragen und gegebenenfalls Mitspracherecht einzuräumen. Nicht zuletzt sollten Evaluierende wissen, wie sich die Datenerhebung in den Lehr-Lern-Prozess einbinden lässt und diesen sogar unterstützen kann. Dieses Wissen sollten die Evaluierenden auch in die Praxis umsetzen können.

**Datenverarbeitung, -auswertung, -interpretation** Die Verarbeitung von Daten geschieht heutzutage vornehmlich computergestützt. Evaluierende sollten daher insbesondere den Einsatz von Office-Software zur Sammlung, Aufbereitung und Auswertung von Daten beherrschen. Office-Software hat gegenüber spezieller Software zur Auswertung von quantitativen oder qualitativen Daten neben meist geringeren Kosten den Vorteil, dass sie in der Regel in den Weiterbildungsorganisationen

bereits vorhanden ist, auf den meisten Computern installiert ist, und kein Experten/innen-Wissen für ihre Anwendung erfordert. Dennoch können Kenntnisse von spezieller Auswertungssoftware für qualitative und/oder quantitative Daten von Vorteil sein. Da in zahlreichen Bildungshäusern eine spezielle Verwaltungssoftware zum Einsatz kommt, ist es auch sinnvoll, deren Funktionsumfang zu kennen, um Integrationsmöglichkeiten in den Evaluationsprozess abschätzen zu können. Bezüglich der Auswertung quantitativer (zum Teil auch qualitativer) Daten sollten Evaluierende in der Lage sein, einfache Häufigkeitsverteilungen anzufertigen. Dabei sollten die Evaluierenden insbesondere mit geringen Fallzahlen operieren und den Umgang mit sinnvollen Prozentuierungen beherrschen. Evaluierende sollten zudem Abweichungen und Extremwerte entsprechend interpretieren und einbeziehen können sowie die Grenzen einer Mittelwertberechnung berücksichtigen können. Eine besondere Rolle spielt in diesem Zusammenhang auch die Übersicht über Möglichkeiten und Grenzen der Datenaggregation, also der Zusammenfassung von Daten auf verschiedenen Ebenen. Auch wenn inferenzstatistische Berechnungen nicht zum Standardbedarf der Organisationen gehören dürften, sollten die Evaluierenden mit probabilistischen Denkweisen vertraut sein. So sollten sie in der Lage sein, einfache Unterschiedshypothesen zu testen, also Vergleiche von Gruppen bezüglich ihrer Mittelwerte und Häufigkeitsverteilungen durchzuführen und Ergebnisse vor dem Hintergrund ihrer Eintrittswahrscheinlichkeit interpretieren können. In jedem Fall sollten Evaluierende Ergebnisse in Tabellen und Grafiken aufbereiten und dabei Qualitätsstandards einhalten können. Bezüglich der Auswertung qualitativer Daten erscheinen Kompetenzen bezüglich der Transkription von Interviews aufgrund des immensen Aufwandes weniger sinnvoll, wohl aber alternative Möglichkeiten der Dokumentation und Zusammenfassung qualitativer Daten. Um Antworten auf offene Fragen in Erhebungsbögen adäquat auswerten zu können, ist für die Evaluierenden die Beherrschung von entsprechenden Auswertungstechniken, wie die Inhaltsanalyse, von besonderer Bedeutung. Wie Daten interpretiert und bewertet werden, ist eine zentrale Frage bei einer Evaluation und erfordert eine grundlegende Kompetenz bei den Evaluierenden in Weiterbildungsorganisationen. So sollten diese in der Lage sein, Daten zu rezipieren (Stichwort: „Datenlesekompetenz“)<sup>86</sup> und sie sollten Leitfragen, Kriterien, Sollwerte und Standards sinnvoll bei der Bewertung einsetzen können. Darüber hinaus sollten die Evaluierenden, darüber reflektieren können, wer

---

86 Koch (2011) legt in diesem Zusammenhang eine Studie zur Datenkompetenz von Lehrkräften an Schulen vor, die sich damit auseinandersetzt, wie Lehrkräfte Rückmeldungen aus Lernstandserhebungen (Vergleichsarbeiten) verstehen.

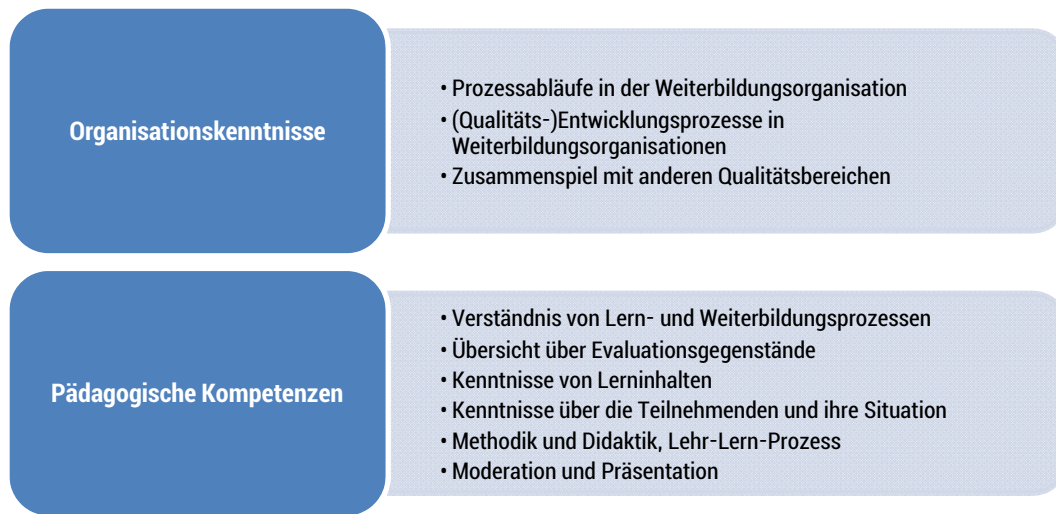
an der Festlegung dieser Bewertungsaspekte beteiligt ist, und wissen welche Standards und Kriterien überhaupt für die Bewertung in Frage kommen. Im Anschluss an die Bewertung müssen die Evaluierenden zudem in der Lage sein, die Ergebnisse so zusammenzustellen, dass sie innerhalb und gegebenenfalls außerhalb der Organisation kommuniziert werden können.

**Projektorganisation** Insbesondere, wenn Evaluationen projektförmig angelegt sind, aber auch wenn Evaluationen als Daueraufgabe in der Organisation verankert sind, erfordert das gesamte Evaluationsprocedere, von der Planung über die Datenerhebung bis zur Datenauswertung und Umsetzung von Konsequenzen, eine personelle und finanzielle Ressourcenplanung sowie eine detaillierte Zeitplanung. Evaluierende in Weiterbildungsorganisationen sollten deshalb fähig sein, den Ressourcen- und Zeitbedarf einzelner Evaluationsschritte abschätzen und kontrollieren zu können. Zur Projektorganisation gehört es ferner, die Organisation der erhobenen Daten zu managen, also fähig zu sein, die Daten zu dokumentieren und zugriffsbereit zu archivieren.

### 7.2.3 Organisationskenntnisse und pädagogische Kompetenzen

Der dritte Kompetenzbereich fasst spezifische Feldkenntnisse von Weiterbildungsorganisationen sowie genuin pädagogische Fähigkeiten zusammen. Aus diesem Grund enthält dieser Bereich im Gegensatz zu dem (zwangsläufig) allgemein gehaltenen Kompetenzfeld „Organisations- und Feldkenntnisse“ des DeGEval-Anforderungsprofils auch sehr viel konkretere Ausführungen. Die Aufnahme der pädagogischen Kompetenzen in diesen Bereich resultiert insbesondere aus dem häufig genannten Hinweis der befragten Organisationsvertreter/innen, dass pädagogische Fähigkeiten unabdingbar für die effiziente Evaluation von Weiterbildungsprozessen seien (vgl. Abschnitt 7.1). Der Kompetenzbereich besteht damit aus zwei Dimensionen a) Organisationskenntnisse und b) Pädagogische Kompetenzen.

**Abb. 51: Dimensionen des Kompetenzbereichs „Organisationskenntnisse und pädagogische Kompetenzen“**



**Organisationskenntnisse** Um überhaupt organisationsrelevante Fragen für die Evaluation bereitstellen, Daten sinnvoll erheben und die Ergebnisse umsetzen zu können, sollten die Evaluierenden detailliert mit den Prozessabläufen in der jeweiligen Weiterbildungsorganisation vertraut sein. Darüber hinaus ist es sinnvoll, wenn die Evaluierenden abschätzen können, welchen Beitrag die Evaluation für die (Qualitäts-)Entwicklungsprozesse leisten kann und wie sie in diese Prozesse integriert werden kann. Dahinter steht gewissermaßen die generelle Kompetenz, die Organisation als lernende und sich entwickelnde Organisation zu verstehen und die Fähigkeit zu haben, dieses Verständnis auch praktisch umzusetzen (vgl. Kapitel 3 zur Lernerorientierten Qualitätstestierung in der Weiterbildung).<sup>87</sup> Wenngleich das LQW-Modell elf Qualitätsbereiche dezidiert unterscheidet, so sind diese Bereiche jedoch nicht hermetisch voneinander getrennt zu betrachten, sondern sie beeinflussen sich gegenseitig, stehen in engem Zusammenhang und bilden eine Einheit. Daher sollten Evaluierende in Weiterbildungsorganisationen, die das LQW-Modell einsetzen, auch das Zusammenspiel der Evaluation mit anderen Qualitätsbereichen berücksichtigen können. Vor allem das Leitbild und die Definition gelungenen Lernens, der Lehr-Lern-Prozess, die Bedarfserschließung und das Controlling erscheinen hier von Bedeutung, wohl aber können auch andere Bereiche, wie etwa die Kundenkommunikation, eng mit der Evaluation von Bildungsprozessen zusammenspielen.

<sup>87</sup> Angesprochen sind hier insbesondere Kenntnisse und Inhalte der Organisations- und Qualitätsentwicklung in Weiterbildungsorganisationen, wie sie beispielsweise in den Grundlagenwerken von Zech u.a. (2006) zu LQW sowie von Zech (2008) zum Qualitätsmanagement in Weiterbildungsorganisationen angesprochen werden. Diese können durch speziellere Kenntnisse der allgemeinen Organisationsentwicklung ergänzt werden, etwa zu Musterwechseln in Organisationen (vgl. von Ameln & Zech, 2011) – um ein mögliches Wissensgebiet zu nennen.

**Pädagogische Kompetenzen** Um Weiterbildungsprozesse evaluieren zu können, sollten die Evaluierenden von selbigen eine klare Vorstellung haben: Wie sieht dieser Weiterbildungsprozess aus? Welche Charakteristika hat er? Wo beginnt und wo endet er? Welche Personen sind daran beteiligt? Unterscheidet er sich von einem (Lehr-)Lern-Prozess und wenn ja, wie? Wo könnten „Knackpunkte“ innerhalb des Prozesses liegen? Diese und viele weitere Fragen sollten die Evaluierenden klären und in der Organisation kommunizieren bzw. gemeinsam erarbeiten können. Das folgende Interviewzitat unterstreicht die Bedeutung dieser Fähigkeit:

„Also ich denke, (...) man muss überhaupt eine Vorstellung davon haben, also ein theoretisches Konstrukt, eine kleine Theorie sage ich mal, dessen was Bildungsprozesse sind. Also das muss man überhaupt erst mal für die Institution geklärt haben. Und zwar auch mit einer gewissen Differenziertheit. Also wenn man sich vorstellt, was hier an der Volkshochschule alles angeboten wird, das sind Kurse der unterschiedlichsten Art und da finden auch äußerst unterschiedliche Prozesse in diesen Kursen statt. Ob man das nun Lernprozesse nennt oder Bildungsprozesse, die unterscheiden sich erheblich voneinander. Und da müsste man erst ein differenziertes Bild dessen haben, was da an Lernprozessen oder eben Bildungsprozessen stattfindet. Das wäre der erste Schritt, also das wäre eine Reflexionsarbeit, wenn man so will.“ (I12:83)

Der Grundgedanke hinter der Klärung der Lern- bzw. Bildungsprozesse ist einfach: Es ist notwendig, die Ziele der Lern- und Bildungsprozesse zu kennen, um Erfolge feststellen zu können. Zudem stellt die Klärung eine optimale Grundlage für die Anwendung von Programmtheorien und logischen Modelle (vgl. hierzu W.K. Kellogg Foundation, 2004, sowie die Beschreibung des Programmbaums in Abschnitt 2.6.3) in der Evaluation dar, wenngleich es sicherlich zu viel verlangt wäre, diese Anwendung als Kompetenz zu fordern. Zu den pädagogischen Kompetenzen gehört weiterhin, sich über die möglichen Evaluationsgegenstände und Evaluationsinhalte in der Weiterbildung im Klaren zu sein und insbesondere die Bedeutung der häufig abgefragten Zufriedenheit in diesem Zusammenhang beurteilen zu können. Evaluierende sollten außerdem Kenntnisse von den Lerninhalten, also fachliches Wissen vom Lerngegenstand, haben, um beispielsweise einen Fragebogen oder ein anderes Evaluationsverfahren konstruieren und den Lernerfolg erfassen zu können. Kenntnisse über die Teilnehmenden und ihre Situation helfen den Evaluierenden, Aussagen von Teilnehmenden zu verstehen, einzuordnen und zu bewerten. Insbesondere im Umgang mit speziellen Zielgruppen, wie etwa Migrantinnen, sollten die Evaluierenden

nicht nur Kenntnis über diese Klientel haben, sondern auch über interkulturelle Kompetenzen verfügen, wie etwa Donna Mertens, die Vertreterin der transformativen Evaluation, fordert (vgl. Weber, 2012, S. 128). Wenn man mithilfe einer Evaluation Lehr-Lern-Prozesse für die Teilnehmenden optimieren und überprüfen möchte, sollte man zudem in der Lage sein, Lehr-Lern-Prozesse zu gestalten, es sind also methodisch-didaktische Kompetenzen notwendig für die flexible Gestaltung von Lernarrangements – darauf weisen zahlreiche Organisationsvertreter/innen hin. Für die Gestaltung von Feedbackrunden und anderen Evaluationsverfahren innerhalb eines Lehr-Lern-Prozesses erscheinen zudem Kompetenzen der Moderation und Präsentation äußerst sinnvoll.

#### 7.2.4 Lern- und Entwicklungskompetenzen

Der vierte Kompetenzbereich bündelt die Kompetenzen, die insbesondere für die Selbstevaluation von Bedeutung sind. Um die Perspektive der Selbstevaluation und ihre innewohnende Entwicklungskomponente stärker zu betonen, wurde der Bereich nicht wie in den DeGEval-Anforderungen mit „Sozial- und Selbstkompetenzen“, sondern mit „Lern- und Entwicklungskompetenzen“ überschrieben. Im Vergleich zu den DeGEval-Anforderungen wurden einzelne Dimensionen zusammengefasst und stärker konkretisiert, so dass der Bereich die drei Dimensionen a) Lernkompetenz, b) Reflexionskompetenz sowie c) Kommunikations- und Kooperationskompetenz umfasst.

Abb. 52: Dimensionen des Kompetenzbereichs „Lern- und Entwicklungskompetenzen“





**Lernkompetenz** Evaluationen bestätigen nicht immer, dass man auf dem richtigen Weg ist und legen zuweilen nahe, Änderungen des Kurses vorzunehmen. Grundvoraussetzung für die Durchführung von (Selbst-)Evaluationen sind deshalb die Bereitschaft und der Wunsch der Evaluierenden, überhaupt etwas an der bisherigen Situation zu ändern. Dieser Veränderungswille wiederum ist die Vorstufe zur Kritikfähigkeit, die meint, dass die Evaluierenden in der Lage sein sollten, Rückmeldungen ernst zu nehmen, Rückmeldungen der Teilnehmenden aushalten zu können und auf die Sache zu beziehen, das Positive zu bemerken, aber zugleich eine kritische Haltung der eigenen Arbeit gegenüber zu wahren.<sup>88</sup> Um durch die Evaluation erkannte Probleme lösen zu können, sollten Evaluierende zudem die Fähigkeit besitzen, Probleme lösungsorientiert zu betrachten und Strategien für die Klärung der Probleme parat zu haben.

**Reflexionskompetenz** Um die Aussagen von Teilnehmenden einordnen und reflektieren zu können, ist es zunächst notwendig, dass sich die Evaluierenden in die Perspektive der Teilnehmenden versetzen und den (Weiter-)Bildungsprozess aus ihrem Blickwinkel betrachten können. Das LQW-Modell spricht hier von der „Lernerorientierung“, um deutlich zu machen, dass der Fokus aller Qualitätsanstrengungen auf die Lernenden gerichtet sein soll, und so erscheint es nur folgerichtig, auch die Übernahme der Lernendenperspektive im Kontext der Evaluation zu fordern. Damit geht diese Anforderung deutlich über die der sozialen Kompetenz der DeGEval hinaus, denn sie ist ja nicht nur auf einzelne Personen beschränkt, sondern ist als Kompetenz der gesamten Organisation, also all ihrer Mitarbeitenden zu sehen. Den Sinn dieser Anforderung verdeutlicht die folgende Antwort eines Organisationsvertreters auf die Frage, welche Kompetenzen man für die effiziente Evaluation von Bildungsprozessen benötigt:

„Eine Offenheit gegenüber den Antworten. Aber eher so persönliche Geschichten glaube ich muss man haben, dass einen das auch wirklich interessiert und man sich auch eben reinversetzt in den anderen. Das würde ich als Hauptkriterium sehen. Um auch die Punkte definieren zu können, die man evaluiert. Ansonsten kommt so was totes raus, wie ‚Waren die Räume schön eingerichtet?‘“  
(I08:56)

---

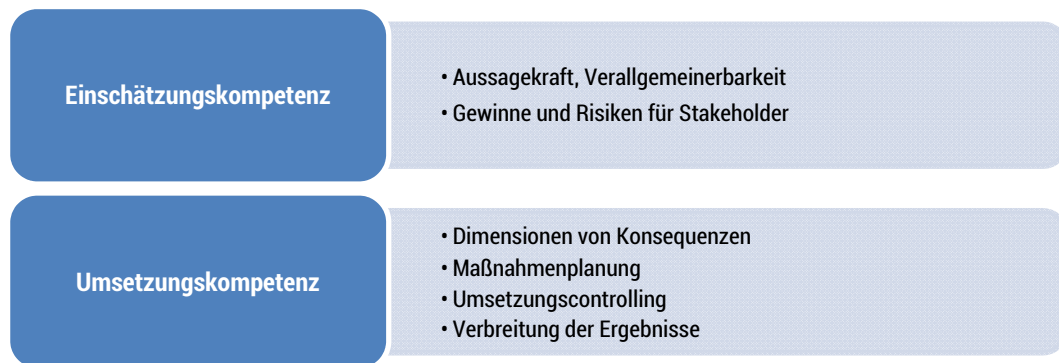
<sup>88</sup> Ansätze zur Erfassung und Konkretisierung dieser Kompetenzen bietet etwa der Erhebungsansatz „The Readiness for Organizational Learning and Evaluation (ROLE)“ von Preskill und Torres (2001).

Zur Reflexionskompetenz gehört zudem die Fähigkeit, Sachverhalte abstrahieren und distanziert betrachten zu können, zumal die Evaluierenden in der Selbstevaluation mit dem evaluierten Gegenstand zum Teil sehr eng verbunden sind und ihre Perspektive eingeschränkt und voreingenommen sein kann. Diese Enge zum Gegenstand erfordert gleichermaßen auch die Kompetenz, die Objektivität wahren zu können.

**Kommunikations- und Kooperationskompetenz** Evaluationsprozesse basieren auf Kommunikation und Kooperation, sei es allgemein zwischen den Mitarbeitenden oder mit den Teilnehmenden oder ganz konkret bei der Rückmeldung der Evaluation an die Lehrenden, wie es das LQW-Modell fordert. Deshalb sollten Evaluierende in Weiterbildungsorganisationen fähig sein, professionelles Feedback zu geben, das Lernen und Entwicklung ermöglicht und keinen Widerstand aufbaut. Zudem erscheinen Fähigkeiten in der Gesprächs- und Verhandlungsführung erforderlich, wenn es beispielsweise darum geht, Evaluationsergebnisse im Team zu besprechen und Konsequenzen zu vereinbaren oder aber auch Interviews mit Teilnehmenden oder anderen Stakeholdern zu führen. Es liegt folgerichtig nahe, auch Konfliktfähigkeit mit in das Anforderungsprofil aufzunehmen, was unter anderem meint, Konflikte erkennen, ansprechen und konstruktiv lösen zu können. Dementsprechend zählen zur Kommunikations- und Kooperationskompetenz, mit Abwehr der Beteiligten und Betroffenen gegenüber der Evaluation, aber auch mit der Belastung der Stakeholder durch die Evaluation umgehen zu können.

### 7.2.5 Nutzungskompetenzen

Der fünfte Kompetenzbereich ist speziell dem Aspekt der Nutzung der Evaluation und ihrer Ergebnisse gewidmet. Zwar greift das Anforderungsprofil der DeGEval diesen Aspekt indirekt auf, indem unter anderem den vier Kompetenzfeldern einzelne Standards zum Nutzen zugeordnet werden (vgl. Caspari et al., 2008, S. 11), doch legt die bisherige Analyse des Status quo und der Herausforderungen in der Evaluationspraxis nahe, den wichtigen Aspekt der Nutzung durch einen eigenen Kompetenzbereich stärker in den Vordergrund zu stellen. Anders als die DeGEval-Anforderungen ist das vorliegende Profil zudem mit Blick auf Selbstevaluierende formuliert, die in der Regel selbst für die Umsetzung der Evaluationsergebnisse verantwortlich sind, was bei externen Evaluierenden nur selten der Fall ist. Die Nutzungskompetenzen umfassen zwei Dimensionen: a) Einschätzungskompetenz und b) Umsetzungskompetenz.

**Abb. 53: Dimensionen des Kompetenzbereichs „Nutzungskompetenzen“**

**Einschätzungskompetenz** Hinter dieser Kompetenz verbergen sich im Wesentlichen zwei Fähigkeiten zur Abschätzung des Nutzens. Zum einen sollten die Evaluierenden in der Lage sein, die Aussagekraft und Verallgemeinerbarkeit von Evaluationsergebnissen einzuschätzen, zum zweiten sollten sie den Gewinn der Ergebnisse und die Folgen möglicher Konsequenzen für die einzelnen Stakeholder, vor allem für die Lernenden, Lehrenden und Planenden, einschätzen können.

**Umsetzungskompetenz** Evaluationen und ihre Ergebnisse führen häufig zu Konsequenzen. Daher erscheint es wichtig, dass die Evaluierenden in Weiterbildungsorganisationen die Ergebnisse und ihre Konsequenzen hinsichtlich verschiedener Dimensionen bewerten können: Welche Relevanz hat das identifizierte Problem? Wie dringend sollte es gelöst werden? Wer ist von der Konsequenz betroffen? Ferner benötigen Evaluierende die Kompetenz, aus den Konsequenzen konkrete, abzuarbeitende Maßnahmen ableiten zu können. Sie sollten also die Maßnahmenplanung beherrschen, womit auch die Fähigkeit zur Steuerung der Maßnahmenumsetzung einhergeht. Optimalerweise achten die Evaluierenden bereits vor der Datenerhebung darauf, dass festgelegt wird, wie bei welchen Ergebnissen gehandelt wird. Schließlich sollten die Evaluierenden fähig sein, die Ergebnisse zu verbreiten und Klarheit darüber schaffen können, wer welche Informationen zu welchem Zeitpunkt und zu welchem Zweck erhält. Für den gesamten Nutzungsprozess bieten insbesondere die im dritten Kompetenzbereich ausgeführten Organisationskenntnisse wertvolle Unterstützung, denn das Wissen über die Prozessabläufe in der Organisation sowie das Wissen über Strategien der Organisationsentwicklung helfen, die Nutzung der Ergebnisse zu planen und umzusetzen.

### 7.3 Übersicht über die Kompetenzbereiche des Anforderungsprofils

Das vollständige Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen ist relativ umfangreich und komplex, weshalb das gesamte Profil abschließend noch einmal in vollständiger Form in Abb. 54 wiedergegeben ist. In der ersten Spalte sind die jeweiligen Dimensionen der fünf Kompetenzbereiche aufgeführt, während die zweite Spalte die jeweils zugehörigen Erläuterungen und Konkretisierungen der Dimensionen enthält.

Abb. 54: Kompetenzbereiche und Dimensionen des Anforderungsprofils in der Übersicht

1. BASISWISSEN	
a) Evaluationsgrundlagen	<ul style="list-style-type: none"> <li>• Evaluationsdefinitionen und Evaluationsverständnis</li> <li>• Funktionen/Zwecke von Evaluationen</li> <li>• Unterscheidungsmerkmale von Evaluationen</li> <li>• Stakeholder</li> </ul>
b) Evaluationsansätze	<ul style="list-style-type: none"> <li>• Ablaufprozesse von Evaluationen</li> <li>• Eignung unterschiedlicher Ansätze und Modelle</li> <li>• Implikationen von relevanten theoretischen und methodologischen Ansätzen und Modellen</li> </ul>
c) Evaluationsstandards	<ul style="list-style-type: none"> <li>• Sicherung der Qualität von Evaluation</li> <li>• Nutzung von Standards</li> </ul>
2. METHODENKOMPETENZEN	
a) Planung und Durchführung empirischer Untersuchungen	<ul style="list-style-type: none"> <li>• Bedeutung, Entwicklung und Beurteilung von Fragestellungen</li> <li>• Steuerung von Umfang, Rhythmus und Zeitpunkten</li> <li>• Berücksichtigung datenschutzrechtlicher Aspekte</li> </ul>
b) Datenerhebung	<ul style="list-style-type: none"> <li>• Möglichkeiten und Grenzen von Erhebungsformen und -instrumenten</li> <li>• Auswahl und Entwicklung von Indikatoren</li> <li>• Entwicklung und Einsatz von Fragebögen sowie Erhebungsinstrumenten jenseits des Fragebogens</li> <li>• Verknüpfung unterschiedlicher Erhebungsformen sowie quantitativer und qualitativer Daten</li> <li>• Kontextabhängigkeit von Datenerhebungen</li> <li>• Konzeption von Abbrecherbefragungen und Nacherhebungen</li> <li>• Einbindung von Lehrenden in die Datenerhebung</li> <li>• Integration in den Lehr-Lern-Prozess</li> </ul>
c) Datenverarbeitung, -auswertung, -interpretation	<ul style="list-style-type: none"> <li>• Einsatz von Office-Software, Integration vorhandener Datenverarbeitungssysteme</li> <li>• Auswertungsverfahren für quantitative und für qualitative Daten</li> <li>• Tabellarische und grafische Aufbereitung von Ergebnissen</li> <li>• Dateninterpretation und Datenbewertung durch Leitfragen, Kriterien, Sollwerte und Standards</li> <li>• Zusammenstellung von Ergebnissen</li> </ul>
d) Projektorganisation	<ul style="list-style-type: none"> <li>• Abschätzung, Einsatz und Kontrolle von Ressourcen</li> <li>• Datenmanagement</li> </ul>

3. ORGANISATIONSKENNTNISSE UND PÄDAGOGISCHE KOMPETENZEN	
a) Organisationskenntnisse	<ul style="list-style-type: none"> <li>• Prozessabläufe in der Weiterbildungsorganisation</li> <li>• (Qualitäts-)Entwicklungsprozesse in Weiterbildungsorganisationen</li> <li>• Zusammenspiel mit anderen Qualitätsbereichen</li> </ul>
b) Pädagogische Kompetenzen	<ul style="list-style-type: none"> <li>• Verständnis von Lern- und Weiterbildungsprozessen</li> <li>• Übersicht über Evaluationsgegenstände</li> <li>• Kenntnisse von Lerninhalten</li> <li>• Kenntnisse über die Teilnehmenden und ihre Situation</li> <li>• Methodik und Didaktik, Lehr-Lern-Prozess</li> <li>• Moderation und Präsentation</li> </ul>
4. LERN- UND ENTWICKLUNGSKOMPETENZEN	
a) Lernkompetenz	<ul style="list-style-type: none"> <li>• Änderungsbereitschaft</li> <li>• Kritikfähigkeit</li> <li>• Problemlösestrategien</li> </ul>
b) Reflexionskompetenz	<ul style="list-style-type: none"> <li>• Lernendenorientierung</li> <li>• Abstraktions- und Distanzfähigkeit</li> <li>• Objektivität</li> </ul>
c) Kommunikations- und Kooperationskompetenz	<ul style="list-style-type: none"> <li>• Feedback</li> <li>• Gesprächs- und Verhandlungsführung</li> <li>• Kooperation und Arbeit in Gruppen</li> <li>• Konfliktfähigkeit</li> <li>• Umgang mit Abwehr und Belastung</li> </ul>
5. NUTZUNGSKOMPETENZEN	
a) Einschätzungskompetenz	<ul style="list-style-type: none"> <li>• Aussagekraft, Verallgemeinerbarkeit</li> <li>• Gewinne und Risiken für Stakeholder</li> </ul>
b) Umsetzungskompetenz	<ul style="list-style-type: none"> <li>• Dimensionen von Konsequenzen</li> <li>• Verbreitung der Ergebnisse</li> <li>• Maßnahmenplanung</li> <li>• Umsetzungscontrolling</li> </ul>

Das Kompetenzraster versteht sich nicht als rein normative Leitlinie, sondern als Reflektions- und Unterstützungsinstrument. In unterschiedlichen Weiterbildungsorganisationen ist von einem organisationsspezifischen Bedarf der Kompetenzen auszugehen. Deshalb sollte bei der Anwendung des Anforderungsprofils zunächst bei jeder Kompetenz bewertet werden, welcher Stellenwert dieser im konkreten Anwendungskontext beigemessen werden soll. Ein solches an die Situation angepasstes Vorgehen trägt auch der Tatsache Rechnung, dass Kompetenzanforderungen nicht gleichermaßen für alle gelten können, sondern ihre Anwendung stark davon abhängt, wer genau in welchem Kontext evaluiert (vgl. Rädiker & Niklesz, 2012).

Das Anspruchsniveau der Kompetenzen bleibt im Bereich des Basiswissens und der Methodenkompetenzen sicherlich hinter den Kompetenzen von externen, speziell geschulten Evaluierenden zurück. In den Bereichen Organisationskenntnisse/päda-

gogische Kompetenzen, Lern- und Entwicklungskompetenzen sowie Nutzungskompetenzen kommt verstärkt die Eigenart der Selbstevaluation zum Tragen. Deshalb liegt das Anforderungsniveau in diesem Bereich deutlich höher als man von externen Evaluierenden erwarten würde, die Weiterbildungsprozesse in einer Organisation evaluieren.

Abschließend sei darauf hingewiesen, dass die Kompetenzbereiche nicht isoliert voneinander zu betrachten sind. Wenngleich die fünf Bereiche klare und unterschiedliche Kompetenzfelder markieren, so bedingen und beeinflussen sie sich doch gegenseitig. Das Wissen über Auswahlverfahren aus dem zweiten Bereich der „Methodenkompetenzen“ ist beispielsweise notwendig, um die Aussagekraft und Verallgemeinerbarkeit von Ergebnissen im fünften Bereich „Nutzungskompetenzen“ einschätzen zu können.

## 8 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden insgesamt drei Fragestellungen mithilfe einer empirischen Studie unter Organisationen, welche als Qualitätsmanagementverfahren die Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW) einsetzen, beantwortet. Erstens wurde herausgearbeitet, wie die Weiterbildungsprozesse in den untersuchten Organisationen evaluiert werden. Zweitens wurde eruiert, welche Herausforderungen und Probleme bei eben dieser Evaluation der Bildungsprozesse auftreten (können). Und drittens wurden die empirischen Ergebnisse und die zuvor dargestellten theoretischen Ausführungen in einem Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen zusammengeführt. In diesem abschließenden Kapitel 8 werden die Ergebnisse zu diesen drei Themenblöcken noch einmal verdichtet präsentiert und vor dem Hintergrund der theoretischen Ausführungen über die Evaluation von Weiterbildungsprozessen und über das LQW-Modell reflektiert. Dabei werden zum Teil auch Ausblicke auf weiterführende Fragestellungen gegeben, die sich im Prozessverlauf herauskristallisiert haben.

### *Fragestellung 1: Status quo der Evaluation von Bildungsprozessen*

Um den Status quo zu untersuchen, wurde eine inhaltlich strukturierende qualitative Inhaltsanalyse von 50 Selbstbeschreibungen der Evaluationspraxis in LQW-Organisationen durchgeführt. Es zeigt sich, dass die Organisationen Evaluation vorwiegend zum Zwecke der Erfolgskontrolle und Zielüberprüfung betreiben. Die Evaluationspraxis steht insofern im Einklang mit Definitionselementen der Evaluation von Bildungsprozessen im LQW-Praxisleitfaden, wo gleich im ersten Satz der Prüfungsaspekt von Evaluation genannt wird: „Evaluation der Bildungsprozesse bedeutet, dass die durchgeführte Bildungsarbeit von der Organisation mit geeigneten Instrumenten überprüft und bewertet wird“ (Zech, 2006, S. 68). Neben den Evaluationszweck der Überprüfung tritt bei den Organisationen vornehmlich das Ziel der Verbesserung, Optimierung und Entwicklung. Größtenteils können in diesem Zusammenhang die Selbstbeschreibungen der Evaluationspraxis so interpretiert werden, dass die Organisationen mit der Evaluation eine *Verbesserung der Bildungsprozesse* anstreben; offen bleibt dabei jedoch meist die Frage, ob es auch um die *Verbesserung von Bildungsprozessen für die Teilnehmenden* geht, also inwieweit im Bereich der Evaluation die von LQW postulierte Lernerorientierung greift und die Evaluation dazu beiträgt, dass die einzelne Organisation ihrer Idealvorstellung gelungenen Lernens in der täglichen Bildungsarbeit näher kommt. Hier bedarf es weiterer For-

schung, um den Zusammenhang zwischen Evaluation und der Lernerorientierung detaillierter zu untersuchen.

Eine erkenntnisgenerierende und eine lernunterstützende Funktion von Evaluation werden nur von wenigen Organisationen explizit genannt. Noch seltener nennen die Organisationen als explizite Evaluationszwecke die Programmplanung und die Bedarfsermittlung, doch geben sie häufig an, die Evaluationsergebnisse hierfür zu verwenden. Bezüglich der evaluierten Gegenstände ist auffällig, dass diese sehr kleinteilig aufgefasst werden und die Grenze zu Bewertungskriterien fließend ist. Dies ist neben anderen Faktoren möglicherweise darauf zurückzuführen, dass die Organisationen ihr Verständnis von Bildungsprozessen nicht klar ausgewiesen haben und in der Folge auch die Gegenstände wenig klar konturiert sind. Die Weiterbildungseinrichtungen setzen als Evaluationsverfahren vorwiegend Fragebögen für Teilnehmende und Lehrende sowie informelle und geplante Gespräche ein, und zwar sowohl im Seminarprozess als auch außerhalb. Eine breite Vielfalt an möglichen Datenquellen und Evaluationsverfahren wird nur von wenigen Organisationen beschrieben. Es lässt sich daher vermuten, dass das Potenzial, das Methoden und Datenquellen jenseits des klassischen Fragebogens wie etwa der Verwendung vorhandener Daten innewohnt, noch nicht ausgeschöpft ist – oder aber bereits umgesetzt wird, sich aber nicht in den analysierten Selbstbeschreibungen widerspiegelt. Die Weiterbildungsorganisationen führen selten Vollerhebungen durch, sondern beschränken ihre Datenerhebung beispielsweise auf einzelne Programm- oder Themenbereiche sowie auf neuartige Veranstaltungen oder Veranstaltungen mit neuen Lehrenden.

Die Prozesse der Datenauswertung, der Datenbewertung und der Ergebnisnutzung verschwimmen häufig in den Selbstbeschreibungen und lassen sich nicht immer klar voneinander abgrenzen. Dabei bleibt insbesondere unklar, wie die Datenauswertung genau durchgeführt wird. Obwohl die Bewertung üblicherweise als das kennzeichnende und unverzichtbare Element einer Evaluation verstanden wird und obwohl auch das LQW-Modell eine Bewertung der Analysen vorschreibt, finden sich bei fast der Hälfte der Selbstreporte keine genauen Angaben zur Bewertung im Rahmen der Evaluation. Die andere Hälfte der Organisationen gibt klare Sollwerte an, die sich meist auf quantitative Daten beziehen und vor deren Hintergrund die Ergebnisse bewertet werden. Die Weiterbildungseinrichtungen verwenden die Evaluationsergebnisse vorwiegend intern und nur wenige informieren die Teilnehmenden und die Öffentlichkeit von den Ergebnissen.



Die Organisationen führen fast ausschließlich interne bzw. Selbstevaluationen durch, kaum eine Einrichtung setzt auf externe Dienstleistungen und wenn, dann häufig nur auf kostenfreie Unterstützung wie etwa von Diplomanden. Auffällig ist zudem, dass in keiner der untersuchten Selbstbeschreibungen Bezug auf Evaluationsansätze oder -modelle genommen wird. Möglicherweise besteht jedoch in der Anwendung von Evaluationsansätzen und -modellen und der Berücksichtigung von deren Implikationen und Denkweisen ein großes Potenzial für die Reflexion und Gestaltung von Evaluationsprozessen in Weiterbildungsorganisationen. Dies wird in dem in Kapitel 7 entwickelten Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen im Detail ausgeführt.

Bemerkenswert ist zudem, dass sich in den Selbstbeschreibungen zur Evaluation der Bildungsprozesse nur wenige Querbezüge zu anderen Qualitätsbereichen des LQW-Modells finden und daher die Evaluation der Bildungsprozesse relativ isoliert dasteht. Gerade bezüglich der Verbindung mit anderen Qualitätsbereichen ist ein großes Potenzial zu vermuten. Dies könnte beispielsweise ausgeschöpft werden, wenn man bei der Evaluation verschiedene Tragweiten und Schnelligkeiten unterscheidet: So ließen sich ad hoc Evaluationen, die meist im Seminargeschehen stattfinden, gut mit dem Qualitätsbereich 4 zu den Lehr-Lern-Prozessen verknüpfen. Evaluationen, welche mittelfristige Fragestellungen verfolgen, etwa bezogen auf ein Semester, ließen sich gut mit Qualitätsbereich 2 zur Bedarfserschließung kombinieren und Evaluationen, welche langfristige Fragestellungen fokussieren, etwa unter einer Fünf-Jahresperspektive, ließen nicht nur ebenfalls gut mit dem Qualitätsbereich 2 zur Bedarfserschließung, sondern auch mit dem Bereich 9 zum Controlling und 11 zu den strategischen Entwicklungszielen verbinden. Entsprechende zeitlich differenzierte Ziele, die als Soll-Werte in die Evaluation einfließen, könnten hierbei eine zusätzliche Unterstützung darstellen.

In der durchgeführten Studie wurden LQW-Organisationen berücksichtigt, die sich einem der folgenden vier Typen zuordnen lassen: (a) sozial/kulturell orientierte und (b) weltanschaulich gebundene Organisationen sowie (c) Volkshochschulen und (d) kommerziell orientierte/wirtschaftsnahe Organisationen. Bei diesen verschiedenen Organisationstypen zeigt sich eine in Details unterschiedliche Evaluationspraxis, beispielsweise setzen kommerziell orientierte bzw. wirtschaftsnahe Organisationen häufiger als andere Organisationen Berichte als Evaluationsverfahren ein. Allerdings lässt sich für die unterschiedlichen Organisationstypen keine grundsätzlich unterschiedliche Evaluationspraxis konstatieren. Der Organisationstyp erklärt

vielmehr nur einen Teil der unterschiedlichen Herangehensweisen und vermutlich spielen auch kulturelle Eigenarten der Organisation, Führungsstile, äußere Rahmenbedingungen und weitere Faktoren eine Rolle.

Es stellt sich die Frage, ob und wenn ja inwieweit die gefundenen empirischen Ergebnisse über den Status quo der Evaluationspraxis den Organisationen eine Unterstützung und Hilfe bei der Weiterentwicklung der organisationsspezifischen Evaluationspraxis bieten können, welchen Nutzen also die Weiterbildungsorganisationen aus den Ergebnissen ziehen können. Für den Zweck der Reflexion der Evaluationspraxis wurde auf Basis der Analyseergebnisse eine Liste mit zahlreichen Fragen erstellt, die es Weiterbildungsorganisationen ermöglicht, einzelne Aspekte ihrer Evaluationstätigkeiten zu hinterfragen (vgl. Abschnitt 5.6). Darüber hinaus könnte es hilfreich sein, die im Internet bereits vorhandenen LQW-Arbeitshilfen zum Thema Evaluation um ausgewählte Inhalte zu ergänzen. Insbesondere könnte sich als nützlich erweisen, Hinweise und Hilfestellungen für die Definition des Begriffs „Bildungsprozess“ zu geben, damit jede Organisation reflektieren kann, was sie genau darunter versteht. Hierfür sprechen drei Gründe: Erstens hat keine der Organisationen in den analysierten Selbstbeschreibungen ihr Verständnis der evaluierten Bildungsprozesse ausgelegt; zweitens finden sich auch in der Literatur zu LQW nur vereinzelte Hinweise, was einen Bildungsprozess kennzeichnet; und drittens benennen einzelne Organisationsvertreter/innen in den durchgeführten Interviews es einerseits als Problem, wenn die Bildungsprozesse einer Organisation nicht klar definiert sind, und andererseits als Kompetenz, wenn Evaluierende es beherrschen, organisationsspezifische Bildungsprozesse zu definieren.

Da in den Beschreibungen häufig die Grenze zwischen den evaluierten Gegenständen und den Bewertungskriterien verschwimmt und weil viele Organisationen keine klaren Angaben über die Bewertungspraxis liefern, könnten Hinweise zu den Unterschieden von Gegenständen und Bewertungskriterien sowie zu Bewertungsmethoden ebenfalls eine nützliche Unterstützung darstellen. Darüber hinaus wären möglicherweise Hinweise auf Evaluationsansätze und Evaluationsmodelle – etwa in Form von Literaturverweisen – sinnvoll. Da die Organisationen primär auf das Erhebungsinstrument Fragebogen setzen, dieser allerdings nicht zwangsläufig für den am häufigsten genannten Zweck der Verbesserung und Optimierung geeignete Daten liefert, könnte auch die Bereitstellung einer Matrix hilfreich sein, die unterschiedlichen Evaluationszwecke jeweils geeignete Verfahren zuordnet. Hierdurch erhielten die Organisationen die Möglichkeit, ihr Bewusstsein darüber zu schärfen, welche Verfahren

zur Erfüllung welcher Zwecke besonders geeignet sind und sie wären in der Lage die Vor- und Nachteile der einzelnen Methoden abschätzen zu können.

Mit Blick auf die häufige Beschränkung der Evaluationsverfahren auf Fragebögen und Gespräche könnte sich weitere Forschung als hilfreich erweisen, um die Bandbreite an sinnvollen und nutzbaren Evaluationsmethoden jenseits und in Kombination mit diesen Instrumenten zu erweitern. Aufgrund der knappen und teilweise gar nicht vorhandenen Darstellungen bezüglich der Datenauswertung und -bewertung, könnte Forschung zudem dazu beitragen, die Gründe hierfür zusammenzutragen.

### *Fragestellung 2: Bei der Evaluation auftretende Herausforderungen*

Um zu untersuchen, welchen Herausforderungen Weiterbildungsorganisationen bei der Evaluation von Bildungsprozessen gegenüberstehen, wurden 24 Interviews mit Personen durchgeführt, die in einer LQW-Organisation für die Evaluation von Bildungsprozessen verantwortlich sind. Die Interviews wurden ebenfalls mithilfe einer inhaltlich strukturierenden qualitativen Inhaltsanalyse ausgewertet. Dabei zeigen sich insgesamt sieben Hauptproblembereiche, die hier nach absteigender Häufigkeit ihrer Nennung aufgelistet sind. Die Herausforderungen bestehen hinsichtlich ...

1. *der Instrumente und Verfahren*, z.B. bezüglich des Messens und Operationalisierens oder bezüglich des Spannungsfelds von Spezialisierung und Vereinheitlichung,
2. *der Datenerhebung*, z.B. durch die Verzerrung von Daten aufgrund von Grundeinstellungen der befragten Teilnehmenden oder aufgrund geringer Rückläufe,
3. *der Akzeptanz und Belastung von Teilnehmenden, Lehrenden und Mitarbeitenden*, z.B. durch mehrfaches Ausfüllen von Bögen durch die Teilnehmenden,
4. *der Datenauswertung und -bewertung*, z.B. aufgrund unterschiedlicher Ansprüche bezüglich Datenaggregation und Detailauswertung,
5. *des Aufwandes und der Ressourcen*, z.B. weil eine große Masse an Daten verarbeitet werden muss oder Evaluation als belastende Zusatzaufgabe wahrgenommen wird,
6. *des Nutzens, des Erkenntnisgewinns und der Aussagekraft*, z.B. weil die Daten keine differenzierenden Erkenntnisse liefern oder die Ergebnisse „zu“ positiv ausfallen sowie
7. *der Nutzung, Umsetzung und Ergebnisverwendung*, z.B. aufgrund von Unklarheiten darüber, wer für die Umsetzung der Konsequenzen verantwortlich ist.

Wie beim Status quo zeigen sich bei den Herausforderungen ebenfalls nur geringe Unterschiede bezüglich verschiedener Organisationstypen und -größen. Auch die Art des Bildungsangebots und die Veranstaltungsformen stehen in keinem bzw. keinem nennenswerten Zusammenhang zu den Herausforderungen. Man kann daraus nicht schließen, dass die Probleme eher organisationsspezifischer Art sind – dafür zeigen sich zu viele deutliche Häufungen und Ähnlichkeiten bei den Hauptproblembereichen. Allerdings lassen sich die Organisationen anhand dieser hauptsächlich vorgetragenen Herausforderungen in vier Gruppen einteilen:

- Organisationen des Typs „Instrumente und Verfahren im Fokus“ (13 Organisationen) stehen vorwiegend vor der Frage „Wie gestalten wir unsere Instrumente und Verfahren?“ An die primär methodischen Probleme schließen sich teilweise nahtlos noch weitere Herausforderungen an, etwa wie man Instrumente gestaltet, die zu einer ertragreichen Nutzung führen.
- Organisationen des Typs „Ergebnisse im Fokus“ (5 Organisationen) stehen vorwiegend vor der Frage „Wie nutzen wir die Evaluationsergebnisse effizient?“ Diese Gruppe bemängelt demnach vorwiegend die Umsetzung der Evaluationsergebnisse und die Ergebnisverwendung.
- Organisationen des Typs „Aufwand im Fokus“ (2 Organisationen) stehen vorwiegend vor der Frage „Wie bewältigen wir (auch noch) die Evaluation?“ Zentral für diese Gruppe sind die Belastung und der Arbeitsaufwand, der durch die Evaluation entsteht.
- Organisationen des Typs „Konzeption im Fokus“ (4 Organisationen) stehen vorwiegend vor der Frage „Wie organisieren und konzeptionieren wir die Evaluation als Gesamtprozess?“ Während die anderen Organisationen eher auf problematische Einzelaspekte hinweisen, steht bei dieser Gruppe die Gesamtsicht auf den Evaluationsprozess als Ganzes im Vordergrund.

Die entwickelte Typologie lässt sich beispielsweise verwenden, um Fortbildungsangebote für Evaluierende in Weiterbildungsorganisationen zu konzipieren, welche auf einzelne Typen zugeschnitten sind und die entsprechenden Problembereiche thematisieren. Dabei gilt es jedoch zu berücksichtigen, dass Organisationen, die sich stark auf einen speziellen Problembereich fokussieren, eventuell auch andere Bereiche ausgeblendet haben. Wenn beispielsweise die Instrumente und Verfahren besonders stark als Probleme betont werden, könnte dies zwar als Anknüpfungspunkt für eine Fortbildung verwendet werden. Die Fortbildung sollte jedoch die Instru-

mente und Verfahren in einen größeren Zusammenhang stellen, etwa Bezüge zu den Evaluationszwecken und Auswertungsaspekten herstellen, weil wahrgenommene Probleme bezüglich der Instrumente auch Ausdruck einer unzureichenden Zielklärung in der Planungsphase einer Evaluation sein könnten.

Die durchgeführte Studie hat nicht nur Fragen bezüglich der Herausforderungen beantwortet, sondern auch tiefergehende aufgeworfen. So wäre es interessant, den Hauptproblembereich der Instrumente und Verfahren mit den Bereichen des Nutzens und der Nutzung zu verbinden und zu erheben, welchen Ertrag eine Einrichtung mit unterschiedlichen Instrumenten und Verfahren erzielt. Dadurch ließe sich herauszufinden, bei welchem minimalen Einsatz, dennoch gut verwertbare Ergebnisse zu erwarten sind. Besonders interessant wäre außerdem, Organisationsvertreter/innen nach ihrem Verständnis von Bildungsprozessen zu befragen, denn in den Interviews scheinen diesbezüglich teilweise sehr unterschiedliche Sichtweisen heraus. Mal wird der Bildungsprozess eng gefasst und mit dem Lehr-Lern-Prozess gleichgesetzt, mal kristallisiert sich ein etwas umfassenderes Verständnis heraus, das auch den Lehr-Lern-Prozess umgebende Aspekte einschließt, und mal wird die Evaluation von Bildungsprozessen explizit gegen die Evaluation von Teilnehmerszufriedenheit abgegrenzt. Eine weitere Erhebung könnte hier zu einem deutlich besseren Verständnis der Denkweise in den Organisationen führen, zumal das Verständnis von Bildungsprozessen auch Auswirkungen auf andere Bereiche jenseits der Evaluation hat.

### *Fragestellung 3: Kompetenzanforderungen an Evaluierende in Weiterbildungsorganisationen*

Unter Berücksichtigung der empirischen Ergebnisse zum Status quo der Evaluationspraxis und den bei der Evaluation der Bildungsprozesse auftretenden Herausforderungen wurde abschließend ein Anforderungsprofil für Evaluierende in Weiterbildungsorganisationen formuliert. Dieses Anforderungsprofil geht von Evaluierenden aus, die Selbst- bzw. interne Evaluation durchführen, und zwar vornehmlich von pädagogisch planenden Mitarbeitenden, die zum Teil auch selbst in der Lehre tätig sind. Evaluation gehört daher nicht zur Haupttätigkeit dieser Personen, sondern die Evaluation der Bildungsprozesse stellt eine Teilaufgabe ihrer pädagogischen Arbeit dar. Als Ausgangsbasis für die Formulierung des Anforderungsprofils diente das seit 2004 unter dem Titel „Empfehlungen für die Aus- und Weiterbildung in der Evaluation – Anforderungsprofile an Evaluatorinnen und Evaluatoren“ vorliegende Kompetenzraster der DeGEval. In das Anforderungsprofil sind explizit auch

die Antworten der befragten Organisationsvertreter/innen auf die Frage, welche Kompetenzen man ihrer Ansicht nach für eine effiziente Evaluation der Bildungsprozesse benötige, eingeflossen. Dem Profil liegt dabei ein Verständnis von Kompetenzen zugrunde, welches sowohl Wissen, Fertigkeiten als auch Haltungen umfasst. Das neu entwickelte Profil enthält fünf Kompetenzfelder, denen jeweils zwei bis vier Kompetenzdimensionen zugeordnet sind. Im Folgenden werden die fünf Kompetenzbereiche kurz vorgestellt:

- *Basiswissen:* Der erste Kompetenzbereich stellt vorwiegend die Wissensgrundlage für die weiteren Felder dar und umfasst insbesondere den Bereich von Evaluationsgrundlagen, wie etwa unterschiedliche Definitionen und Unterscheidungsmerkmale von Evaluationen. Ferner gehört auch das Wissen um Evaluationsansätze und deren Implikationen für die Praxis sowie die Berücksichtigung von Evaluationsstandards zum ersten Kompetenzfeld.
- *Methodenkompetenzen:* Der zweite Kompetenzbereich stellt den Schwerpunkt des Profils dar und umfasst zunächst die Planung und Durchführung empirischer Untersuchungen, wie etwa die Entwicklung und Verwendung geeigneter Fragestellungen. Ferner gehören auch die Datenerhebung sowie die sich anschließende Datenverarbeitung, -auswertung und -interpretation zu diesem Bereich und beinhaltet unter anderem die Fähigkeit zur Auswahl und Entwicklung von Indikatoren wie auch zur Bewertung der Ergebnisse anhand von Leitfragen. Die Fähigkeit der Projektorganisation zählt ebenfalls dazu.
- *Organisationskenntnisse und pädagogische Kompetenzen:* In diesem Bereich sind Organisationskenntnisse, etwa über Prozessabläufe und Qualitätsentwicklungsmaßnahmen in der Organisation, sowie pädagogische Kompetenzen zusammengefasst. Zu letzteren zählen beispielsweise das weiter oben angesprochene Verständnis von Lern- und Weiterbildungsprozessen, aber auch die Kenntnisse über die Teilnehmenden und deren Situation.
- *Lern- und Entwicklungskompetenzen:* Dieser Kompetenzbereich ergibt sich aus der speziellen Tatsache, dass das Personal Selbstevaluation der eigenen Arbeit betreibt und daher über Lernkompetenz, Reflexionskompetenz sowie Kommunikations- und Kooperationskompetenz verfügen sollte. Dies versetzt das Personal in die Lage, sich und die Organisation weiterzuentwickeln.
- *Nutzungskompetenzen:* Auch der fünfte und letzte Kompetenzbereich trägt der Tatsache Rechnung, dass es sich um Selbstevaluation handelt. In diesen Bereich

fallen sowohl die Kompetenz, die Aussagekraft und Verallgemeinerbarkeit von Ergebnissen einzuschätzen, als auch die Umsetzungskompetenz, die dazu befähigt, adäquate Maßnahmen zu planen und umzusetzen.

Das entwickelte Anforderungsprofil berücksichtigt zwar an einigen Stellen die Charakteristika des LQW-Modells, ist aber problemlos auch auf Weiterbildungsorganisationen übertragbar, welche die Lernerorientierte Qualitätstestierung nicht einsetzen. Es lässt sich sowohl für die Entwicklung von Curricula und Fortbildungskonzepten als auch als Reflexionsinstrument zur Personalentwicklung verwenden. Abschließend stellt sich noch die Frage, wie Evaluierende in Weiterbildungsorganisationen diese als sinnvoll betrachteten Kompetenzen im Sinne eines „Evaluation Capacity Building“ erwerben können. Einige Interviewpersonen haben hierzu erste Hinweise gegeben und betonen insbesondere das Erfahrungslernen im Sinne von „Learning-by-Doing“. Auch die Selbstbeschreibungen zeugen an mehreren Stellen von einem Kompetenzentwicklungsprozess, der durch die Einführung und Anwendung des LQW-Ansatzes angestoßen wird und zumeist darin besteht, dass zunächst ohne Blick auf die spätere Nutzung umfangreiche Datenerhebungen stattfinden, die im Laufe der Jahre reduziert und fokussiert werden (vgl. auch Rädiker, 2012, S. 48). Offen bleibt dennoch, inwieweit eine Fortbildung oder auch ein Selbststudium mithilfe einschlägiger Praxisliteratur wie etwa der Bücher von Reischmann (2006) oder von Beywl et al. (2007) geeignet wären, zum Aufbau von Evaluation Capacity beizutragen. Hierzu bedarf es weiterer Forschung, wobei ein erster Schritt wäre, das entwickelte Anforderungsprofil an die Organisationen im Sinne einer kommunikativen Validierung (Mayring, 2002, S. 147) zurückzuspiegeln und zu überprüfen.





## Danksagung

Bei der Erstellung dieser Arbeit konnte ich auf zahlreiche Unterstützung bauen, für die ich mich an dieser Stelle herzlich bedanke.

Mein besonderer Dank gilt Prof. Dr. Udo Kuckartz, dem Betreuer dieser Arbeit, der mir zu jeder Zeit mit einem offenen Ohr für meine Fragen zur Seite stand und zum Gelingen des Werkes durch zahlreiche wertvolle Rückmeldungen zum Inhalt und zum strategischen Vorgehen beigetragen hat. Darüber hinaus danke ich Prof. Dr. Kuckartz für das in mich gesetzte Vertrauen und die zahlreiche zusätzliche Unterstützung über die zahlreichen Rückmeldungen hinaus.

Bei Prof. Dr. Rainer Zech, dem Zweitgutachter dieser Arbeit, bedanke ich mich insbesondere für die große Unterstützung bei der Datenbeschaffung (ohne die ein großer Teil dieser Arbeit nicht realisierbar gewesen wäre!) sowie für die zahlreichen kleinen und große Hinweise, welche diese Arbeit von Beginn an bereichert haben.

Mein Dank geht ebenfalls an alle Interviewpartner und Interviewpartnerinnen, die im Rahmen der durchgeführten Studie bereit waren, ihre kostbare Zeit zur Verfügung zu stellen und meine Fragen zu beantworten. Julia Schehl spreche ich meinen Dank für die zügige und enthusiastische Transkription der Interviews aus.

Julia Busch hat im Rahmen ihrer Diplomarbeit zahlreiche der in dieser Arbeit ausgewerteten Interviews geführt. Sie stand als Diskussionspartnerin bei der Instrumententwicklung zur Verfügung und hat große Teile der Codierarbeit übernommen – vielen Dank an dieser Stelle für diese Unterstützung.

Patrick Plettenberg gebührt mein spezieller Dank für das Ausreizen der Funktionalität des Literaturverwaltungsprogramms Endnote, der Kreativität bei der Erstellung und Überarbeitung von Grafiken sowie bei der Unterstützung der Codierarbeit. Auch Stefanie Zanetti sei herzlich für die Codierunterstützung gedankt. Ein großes Dankeschön geht an Thomas Ebert, Heiko Grunenberg, Lena Lehmann, Julia Schehl sowie Mailin Gunkel und Dina Weiler für ihre Rückmeldungen zu einzelnen Kapiteln.

Schließlich danke ich ganz herzlich meiner Frau Marina Rädiker für die hilfreichen Kommentare zum Manuskript, vor allem aber dafür, dass sie mir mit großem Zeiteinsatz den Rücken für die Fertigstellung dieser Arbeit freigehalten hat.



## Literaturverzeichnis

- Alliger, G. M. & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. *Personnel Psychology*, 42(2), 331-342.
- Ambos, I., Egetenmeyer, R. & Scheller, F. (2008). Personalentwicklung und wirtschaftliches Klima bei Weiterbildungsanbietern. Zentrale Ergebnisse im Überblick. *wbmonitor*. Verfügbar unter: [https://www.wbmonitor.de/downloads/Ergebnisse\\_20090226.pdf](https://www.wbmonitor.de/downloads/Ergebnisse_20090226.pdf) [04.04.12].
- American Evaluation Association. (1995). Guiding Principles for Evaluators. *New Directions for Program Evaluation*, 66(Summer 1995), 19-26.
- Arbeitsstab Forum Bildung in der Geschäftsstelle der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung. (2001). *Qualitätsentwicklung und Qualitätssicherung im internationalen Wettbewerb. Vorläufige Empfehlungen des Forum Bildung. Band 8*. Verfügbar unter: <http://www.blk-bonn.de/papers/forum-bildung/band08.pdf> [23.03.12].
- Arnold, R. (1994). Qualitätssicherung in der Weiterbildung. *Grundlagen der Weiterbildung*, 5(1), 6-10.
- Arnold, R., Nolda, S. & Nuissl, E. (2001). *Wörterbuch Erwachsenenpädagogik*. Bad Heilbrunn: Klinkhardt.
- ArtSet. (2006). *Bildung, Lernorientierung und Qualitätsentwicklung. Einführung in Logik und Verfahren von LQW*. Verfügbar unter: [http://www.artset-lqw.de/cms/fileadmin/user\\_upload/Service\\_Allgemeines/Einfuehrung\\_LQW\\_Modellversion\\_3\\_Folien\\_in\\_Word.pdf](http://www.artset-lqw.de/cms/fileadmin/user_upload/Service_Allgemeines/Einfuehrung_LQW_Modellversion_3_Folien_in_Word.pdf) [04.04.12].
- ArtSet. (2012). *Humanität ist effizient. 25 Jahre ArtSet® – ein Portrait*. Hannover: Expressum.
- Balk, M. (2000). *Evaluation von Lehrveranstaltungen. Die Wirkung von Evaluationsrückmeldung*. Frankfurt am Main u.a.: Lang. Zugleich: Dissertation, Universität München, 2000.
- Balzer, L. (2005). *Wie werden Evaluationsprojekte erfolgreich? Ein integrierender theoretischer Ansatz und eine empirische Studie zum Evaluationsprozess*. Landau: Verlag Empirische Pädagogik. Zugleich: Dissertation, Universität Koblenz-Landau, 2005.
- Balzer, L., Frey, A. & Nenniger, P. (1999). Was ist und wie funktioniert Evaluation? *Empirische Pädagogik*, 13(4), 393-413.
- Beywl, W. (2001). Evaluation in der Jugendverbandsarbeit einführen – Hindernisse und Lösungsansätze. In W. Beywl, R. Mecklenburg, J. Richard, T. Schneid & M. Wonik (Hrsg.), *Evaluation im Alltag. Jugendverbände untersuchen ihre Wirkungen* (S. 17-27). Münster: Votum.
- Beywl, W. (2006). Evaluationsstandards als Grundlage für die Aus- und Weiterbildung in Evaluation – Anforderungsprofile für Evaluatoreninnen im Vergleich. In JCSEE, Joint Committee on Standards for Educational Evaluation & J. R. Sanders (Hrsg.), *Handbuch der Evaluationsstandards. Die Standards des „Joint Committee on Standards for Educational Evaluation“*. Übersetzt und für die deutsche Ausgabe erweitert von Wolfgang Beywl und Thomas Widmer (3., erw. und aktual. Aufl., S. 313-336). Wiesbaden: VS Verlag.
- Beywl, W. (2009). Rezension zu Kirkpatrick, Donald L./Kirkpatrick, James D. (2006): Evaluating training programs: The four levels. *Zeitschrift für Evaluation*, 8(1), 127-130.
- Beywl, W. (2010a). *Die Standards für Programmevaluationen des Joint Committee on Standards for Educational Evaluation (JCSEE). Kurzfassung der 3. Auflage 2010*. Verfügbar unter: <http://web.fhnw.ch/ph/tagungen/gute-evaluationen-fur-das-bildungswesen/bilder/ubersetzte-kurzfassung-der-standards.pdf> [12.04.12].
- Beywl, W. (2010b). *Systematisches Bewerten – ein erster Blick*. Paper vorgestellt auf der DeGEval-Jahrestagung 2010, Luxemburg. Verfügbar unter: <http://www.alt.degeval.de/calimero/tools/proxy.php?id=22969> [12.04.12].
- Beywl, W. & Balzer, L. (2008). Die Weichen richtig stellen. Weiterbildungsevaluation – zehn verbreitete Irrtümer und ihre Korrektur. *Weiterbildung, Zeitschrift für Grundlagen, Praxis und Trends*, 19(4), 8-11.
- Beywl, W., Bestvater, H. & Friedrich, V. (2011). *Selbstevaluation in der Lehre. Ein Wegweiser für sichtbares Lernen und besseres Lehren*. Münster u.a.: Waxmann.
- Beywl, W., Kehr, J., Mäder, S. & Niestroj, M. (2007). *Evaluation Schritt für Schritt: Planung von Evaluationen*. Münster: hiba-Verlag.
- Beywl, W., Mecklenburg, R., Richard, J., Schneid, T. & Wonik, M. (2001). Empfehlungen für die Evaluation in der Jugendhilfe. In dies. (Hrsg.), *Evaluation im Alltag. Jugendverbände untersuchen ihre Wirkungen* (S. 28-33). Münster: Votum.
- Beywl, W. & Niestroj, M. (2009). *Das A-B-C der wirkungsorientierten Evaluation*. Köln: Univation. Verfügbar unter: [http://www.eval-wiki.org/glossar/Eval-Wiki:\\_Glossar\\_der\\_Evaluation](http://www.eval-wiki.org/glossar/Eval-Wiki:_Glossar_der_Evaluation) [04.04.12].
- Bihler, W. (2006). *Weiterbildungserfolg in betrieblichen Lehrveranstaltungen. Messung und Einflussfaktoren im Bereich Finance & Controlling*. Wiesbaden: Deutscher Universitäts-Verlag. Zugleich: Dissertation, Universität Hohenheim, 2006.
- Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler* (4., überarb. Aufl.). Berlin u.a.: Springer.
- Brandt, T. (2002). *Qualifikationsanforderungen für Evaluatoren. Überlegungen zur Entwicklung eines Ergänzungsstudiums Evaluation*. Diplomarbeit, Freie Universität Berlin.

- Brandt, T. (2009). *Evaluation in Deutschland. Professionalisierungsstand und -perspektiven*. Münster u.a.: Waxmann. Zugleich: Dissertation, Universität des Saarlandes, 2008.
- Bronner, R. & Schröder, W. (1983). *Weiterbildungserfolg. Modelle und Beispiele systematischer Erfolgssteuerung*. München u.a.: Hanser.
- Brüggemann, H. & Bremer, P. (2005). *Grundlagen Qualitätsmanagement. Von den Werkzeugen über Methoden zum TQM*. Berlin u.a.: Springer.
- Brunner, F. J. & Wagner, K. W. (2008). *Taschenbuch Qualitätsmanagement: Leitfaden für Studium und Praxis* (4., überarb. Aufl.). München: Hanser.
- Buddenberg, V., Holzapfel, C., Uemminghaus, M., Kaiser, A., Hohenstein, K. & Wolter, M. (2007). Beliebter als ihr Ruf. Lernerfolgskontrollen in der Erwachsenenbildung. *DIE-Zeitschrift für Erwachsenenbildung*, 2007(1).
- Buiskool, B. J., Broek, S. D., Lakerveld, J. A. v., Zarifis, G. K. & Osborne, M. (2010). *Key competences for adult learning professionals*. Verfügbar unter: <http://ec.europa.eu/education/more-information/doc/2010/keycomp.pdf> [04.04.12].
- Busch, J. K. (2009). *Evaluation in Weiterbildungseinrichtungen. Eine qualitative Studie zur Verwendung von Evaluationsergebnissen*. Diplomarbeit, Philipps-Universität Marburg.
- Canadian Evaluation Society. (1999). *Essential skills series*. Verfügbar unter: <http://www.evaluationcanada.ca> [22.12.10].
- Canadian Evaluation Society. (2010). *Competencies for Canadian Evaluation Practice*. Verfügbar unter: [http://www.evaluationcanada.ca/txt/2\\_competencies\\_cdn\\_evaluation\\_practice.pdf](http://www.evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice.pdf) [19.01.10].
- Caspari, A., Hennen, M., Scheffler, D., Schmidt, U. & Schwab, O. (2004). *Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile für Evaluatorinnen und Evaluatoren*. Alfter.
- Caspari, A., Hennen, M., Scheffler, D., Schmidt, U. & Schwab, O. (2008). *Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren* (2., unveränd. Aufl.). Mainz. Verfügbar unter: <http://www.alt.degeval.de/calimero/tools/proxy.php?id=24057> [18.04.12].
- DeGEval. (2004). *Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation*. Alfter: DeGEval. Verfügbar unter: [www.alt.degeval.de/calimero/tools/proxy.php?id=24059](http://www.alt.degeval.de/calimero/tools/proxy.php?id=24059) [04.04.12].
- DeGEval. (2008). *Standards für Evaluation* (4., unveränd. Aufl.). Mainz: DeGEval. Verfügbar unter: <http://www.degeval.de/calimero/tools/proxy.php?id=19076> [06.01.11].
- DeGEval. (2011). *Methoden der Evaluation. Positionspapier der DeGEval*. Verfügbar unter: <http://www.degeval.de/publikationen/positionspapiere/item/564-methoden-der-evaluation-positionspapier-03-der-degeval> [12.01.11].
- Deutscher Bildungsrat. (1970). *Strukturplan für das Bildungswesen: verabschiedet auf der 27. Sitzung der Bildungskommission am 13. Februar 1970*. Bonn.
- Diekmann, A. (2007). *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen* (18. Aufl.). Reinbek bei Hamburg: Rowohlt.
- EuropeAid. (2005). *Evaluation – Guidelines – Glossary*. Verfügbar unter: [http://ec.europa.eu/europeaid/evaluation/methodology/glossary/glo\\_en.htm](http://ec.europa.eu/europeaid/evaluation/methodology/glossary/glo_en.htm) [10.01.09].
- European Evaluation Society. (2009). *Results of a Society Wide Consultation on a Proposed Evaluation Competencies Framework for Europe*. Verfügbar unter: <http://www.europeanevaluation.org/images/file/About%20EES/survey%20results.pdf> [04.04.12].
- Faulstich, P. & Zeuner, C. (2010). *Erwachsenenbildung*. Weinheim, Basel: Beltz.
- Flick, U. (2006). *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzung*. Reinbek bei Hamburg: Rowohlt.
- Flick, U. (2007). Zur Qualität qualitativer Forschung – Diskurse und Ansätze. In U. Kuckartz, H. Grunenberg & T. Dresing (Hrsg.), *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis* (2., überarb. und erw. Aufl., S. 188-209). Wiesbaden: VS Verlag.
- Flick, U. (2009). *Sozialforschung: Methoden und Anwendungen; ein Überblick für die BA-Studiengänge*. Reinbek bei Hamburg: Rowohlt.
- Flick, U. (2011). *Triangulation. Eine Einführung* (3., aktual. Aufl.). Wiesbaden: VS Verlag.
- Frankenreiter, J. J. (1996). *Möglichkeiten und Grenzen der indirekten prozessbezogenen Erfolgsbeurteilung von betrieblichen Weiterbildungsmaßnahmen. Ergebnisse einer empirischen Untersuchung*. Darmstadt: Dissertations Druck Darmstadt. Zugleich: Dissertation, Universität Stuttgart, 1995.
- Früh, W. (2001). Kategorienexploration bei der Inhaltsanalyse. Basiswissengeleitete offene Kategorienfindung. In W. Wirth & E. Lauf (Hrsg.), *Inhaltsanalyse: Perspektiven, Probleme, Potentiale* (S. 117-139). Köln: Halem.
- Gerl, H. (1983). Evaluation in Lernsituationen – Ein Beitrag zu reflexivem Lernen. In H. Gerl & K. Pehl (Hrsg.), *Evaluation in der Erwachsenenbildung* (S. 15-92). Bad Heilbrunn: Klinkhardt.
- Gessler, M. & Sebe-Opfermann, A. (2011). Der Mythos „Wirkungskette“ in der Weiterbildung. Empirische Prüfung der Wirkungsannahmen im „Four Levels Evaluation Model“ von Donald Kirkpatrick. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107(2), 270-279.
- Ghere, G., King, J. A., Stevahn, L. & Minnema, J. (2006). A Professional Development Unit for Reflecting on Program Evaluator Competencies. *American Journal of Evaluation*, 27(1), 108-123.

- Gieseke, W. (2002). Evaluation in Zeiten der Veränderung. Ein Essay. *Literatur- und Forschungsreport Weiterbildung*, 2002(50), 68-76.
- Glass, G. V. (1972). Die Entwicklung einer Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München: R. Piper & Co.
- Gnahs, D. (2007). *Kompetenzen – Erwerb, Erfassung, Instrumente*. Bielefeld: W. Bertelsmann.
- Gnahs, D. (2010). Berichts- und Informationssysteme zur Weiterbildung und zum Lernen Erwachsener. In R. Tippelt & A. von Hippel (Hrsg.), *Handbuch Erwachsenenbildung/Weiterbildung* (4., durchgesehene Aufl., S. 279-292). Wiesbaden: VS Verlag.
- Gollwitzer, M. & Jäger, R. S. (2007). *Evaluation. Workbook*. Weinheim u.a.: Beltz PVU.
- Götz, K. (2001). *Zur Evaluierung betrieblicher Weiterbildung* (4., verbess. Aufl., Bd. 1: Theoretische Grundlagen). München: Hampp.
- Grochowiak, K. & Castella, J. (2001). *Systemdynamische Organisationsberatung: Die Übertragung der Methode Hellingers auf Organisationen und Unternehmen. Ein Handlungsleitfaden für Unternehmensberater und Trainer*. Heidelberg: Carl-Auer-Systeme.
- Grunenberg, H. (2001). *Die Qualität qualitativer Forschung. Eine Metaanalyse erziehungs- und sozialwissenschaftlicher Forschungsarbeiten*. Diplomarbeit, Philipps-Universität Marburg. Verfügbar unter: <http://www.g-berg.de/forschung.htm> [18.10.11].
- Grunenberg, H. & Kuckartz, U. (2010). Deskriptive Statistik in der qualitativen Sozialforschung. In B. Friebertshäuser, A. Langer & A. Prengel (Hrsg.), *Handbuch Qualitative Forschungsmethoden in der Erziehungswissenschaft* (S. 487-500). Weinheim: Juventa.
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth Generation Evaluation*. Newbury Park, CA u.a.: Sage.
- Gutknecht-Gmeiner, M. (2009). Evaluation (in) der Erwachsenenbildung. Eine kritische Würdigung der aktuellen Praxis und Analyse möglicher Handlungsfelder. *MAGAZIN erwachsenenbildung.at. Das Fachmedium für Forschung, Praxis und Diskurs*, 2009(7/8), 13-01 - 13-14. Verfügbar unter: <http://erwachsenenbildung.at/magazin/09-7u8/meb09-7u8.pdf>
- Gutknecht-Gmeiner, M., Hense, J. U., Kihm, S., Lüth, K. & Steckhan, H. (2011). *Didaktik der Aus- und Weiterbildung in der Evaluation. Ein Empfehlungspapier des Arbeitskreises Aus- und Weiterbildung in der Evaluation in der Gesellschaft für Evaluation – DeGEval e.V.* Mainz. Verfügbar unter: [http://www.degeval.de/images/stories/Arbeitskreise/AK\\_AUWE/DeGEval\\_AKAUWE\\_Empfehlungen\\_Didaktik.pdf](http://www.degeval.de/images/stories/Arbeitskreise/AK_AUWE/DeGEval_AKAUWE_Empfehlungen_Didaktik.pdf) [07.04.12]
- Gutknecht-Gmeiner, M. & Schlögl, P. (2007). Stand der Qualitätssicherung und -entwicklung bei österreichischen Institutionen der Erwachsenenbildung. In Bundesministerium für Unterricht, Kunst und Kultur (Hrsg.), *Qualitätsentwicklung und -sicherung in der Erwachsenenbildung in Österreich – Wohin geht der Weg? Darstellung der Ergebnisse des Projektes „Instrumente zur Sicherung der Qualität und Transparenz in der Erwachsenenbildung in Österreich“* (S. 31-54). Wien: Bundesministerium für Unterricht, Kunst und Kultur.
- Habeck, S. & Seitter, W. (2012). Das evaluative Anregungspotenzial von Absolventenstudien. Qualität und Nachhaltigkeit der Ausbildung an der Deutschen Blindenstudienanstalt e.V. (blista) in der retrospektiven Sicht von Absolventinnen und Absolventen. In U. Kuckartz & S. Rädiker (Hrsg.), *Erziehungswissenschaftliche Evaluationspraxis* (S. 52-64). Weinheim: Beltz Juventa.
- Häder, M. (2010). *Empirische Sozialforschung: Eine Einführung* (2. Aufl.). Wiesbaden: VS Verlag.
- Hanke, U. (2005). *Evaluieren von Weiterbildung. Theoretische Grundlagen und praktische Vorschläge*. Hamburg: Kovac.
- Häring, K. (2003). *Evaluation der Weiterbildung von Führungskräften. Anspruch und Realität des Effektivitätscontrolling in deutschen Unternehmen*. Wiesbaden: Deutscher Universitäts-Verlag. Zugleich: Dissertation, Universität Augsburg, 2003.
- Hartz, S., Goeze, A. & Schrader, J. (2007). *Empirische Befunde zur Implementierung von LQW 2 in das System der Weiterbildung. Abschlussbericht des Lehrstuhls Erwachsenenbildung/Weiterbildung der Eberhard Karls Universität Tübingen*. Verfügbar unter: <http://www.die-bonn.de/doks/hartz0702.pdf> [04.04.12].
- Hense, J. U. (2006). *Selbstevaluation – Erfolgsfaktoren und Wirkungen eines Ansatzes zur selbstbestimmten Qualitätsentwicklung im schulischen Bereich*. Frankfurt am Main u.a.: Lang. Zugleich: Dissertation, Universität München, 2005.
- Hense, J. U. (2012). *Eröffnungsbeitrag*. Paper vorgestellt auf der Tagung „Forschung über Evaluation: Theorie und Empirie zu Bedingungen, Prozessen, Einflüssen und Wirkungen der Evaluation“, Berlin. Verfügbar unter: [http://www.degeval.de/images/stories/Arbeitskreise/AK\\_BER\\_BILD/Hense\\_-\\_Forschung\\_ueber\\_Evaluation\\_Eroeffnung.pdf](http://www.degeval.de/images/stories/Arbeitskreise/AK_BER_BILD/Hense_-_Forschung_ueber_Evaluation_Eroeffnung.pdf) [31.05.12].
- Hense, J. U. & Mandl, H. (2009). Evaluations- und Selbstevaluationskompetenz von Lehrenden. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität – Bedingungen, Genese, Wirkungen und Messung* (S. 129-140). Weinheim u.a.: Beltz.
- Hermanns, H. (2000). Interviewen als Tätigkeit. In U. Flick, E. von Kardoff & I. Steinke (Hrsg.), *Qualitative Forschung. Ein Handbuch* (S. 360-368). Reinbek bei Hamburg: Rowohlt.
- Hirschauer, S. (2006). Wie geht Bewerten? – Zu einer anderen Evaluationsforschung. In U. Flick (Hrsg.), *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. (S. 405-423). Reinbek bei Hamburg: Rowohlt.

- Holla, B. (2002). *Qualitätsentwicklung in der Weiterbildung durch praxisorientierte Evaluation*. Frankfurt am Main u.a.: Lang. Zugleich: Dissertation, Universität Hildesheim, 2000 u.d.T.: Entwicklung durch praxisorientierte Evaluation.
- Hopf, C. & Schmidt, C. (1993). Zum Verhältnis von innerfamiliären sozialen Erfahrungen, Persönlichkeitsentwicklung und politischen Orientierungen. Hildesheim: Institut für Sozialwissenschaften der Universität Hildesheim. Verfügbar unter: <http://w2.wa.uni-hannover.de/mes/berichte/rex93.htm> [04.04.12].
- JCSEE, Joint Committee on Standards for Educational Evaluation & Sanders, J. R. (2006). *Handbuch der Evaluationsstandards. Die Standards des „Joint Committee on Standards for Educational Evaluation“*. Übersetzt und für die deutsche Ausgabe erweitert von Wolfgang Beywl und Thomas Widmer (3., erw. und aktual. Aufl.). Wiesbaden: VS Verlag.
- Käpplinger, B. (2009). Bildungscontrolling: Vor allem in Großbetrieben ein Thema. BIBB-Umfragen von 1997 und 2008 im Vergleich. *BIBB REPORT*, 13(9), 1-8. Verfügbar unter: [http://www.bibb.de/dokumente/pdf/a12\\_bibbreport\\_2009\\_13.pdf](http://www.bibb.de/dokumente/pdf/a12_bibbreport_2009_13.pdf) [18.04.12].
- Kelle, U. & Kluge, S. (2010). *Vom Einzelfall zum Typus*. Wiesbaden: VS Verlag.
- Kellner, H. J. (2006). *Value of Investment. Neue Evaluierungsmethoden für Personalentwicklung und Bildungscontrolling*. Offenbach: GABAL.
- King, J. A., Minnema, J., Ghore, G. & Stevahn, L. (1998). *Essential Competencies for Program Evaluators*. Paper vorgestellt auf dem Annual Meeting of the American Evaluation Association, Chicago, IL. Verfügbar unter: <http://www.eval.org/SummerInstitute07/Handouts/si07.kingF.pdf> [12.04.12].
- King, J. A., Stevahn, L., Ghore, G. & Minnema, J. (2001). Toward a Taxonomy of Essential Evaluator Competencies. *American Journal of Evaluation*, 22(2), 229-247.
- Kirkpatrick, D. L. & Kirkpatrick, J. D. (2006). *Evaluating training programs. The four levels* (3rd ed.). San Francisco: Berrett-Koehler.
- Koch, U. (2011). *Verstehen Lehrkräfte Rückmeldungen aus Vergleichsarbeiten? Datenkompetenz von Lehrkräften und die Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten*. Münster u.a.: Waxmann. Zugleich: Dissertation, Universität Koblenz-Landau, 2010 u.d.T.: Datenkompetenz von Lehrkräften und die Nutzung von Ergebnissrückmeldungen aus Vergleichsarbeiten.
- Körber, K. (2002). Institution – Organisationen – Anbieter. Anmerkungen zur erwachsenenpädagogischen Organisationsforschung. In I. Ambos & E. Nuissl (Hrsg.), *DIE-Workshop Forschung zur Erwachsenenbildung 11.-13. Januar 2001 in der Evangelischen Akademie Hofgeismar* (S. 231-251). Frankfurt am Main: DIE.
- Krekel, E. M. & Seusing, B. (1999). *Bildungscontrolling. Ein Konzept zur Optimierung der betrieblichen Weiterbildungsarbeit*. Bielefeld: Bertelsmann.
- Kromrey, H. (2001a). Evaluation – ein vielschichtiges Konzept. Begriff und Methodik von Evaluierung und Evaluationsforschung. Empfehlungen für die Praxis. *Sozialwissenschaften und Berufspraxis*, 24(2), 105-131.
- Kromrey, H. (2001b). Studierendenbefragungen als Evaluation der Lehre? Anforderungen an Methodik und Design. In U. Engel (Hrsg.), *Hochschul-Ranking. Zur Qualitätsbewertung von Studium und Lehre* (S. 11-48). Frankfurt am Main: Campus.
- Kromrey, H. (2003). Evaluation in Wissenschaft und Gesellschaft. *Zeitschrift für Evaluation*, 2(1), 93-116.
- Kromrey, H. (2006). *Empirische Sozialforschung. Modelle und Methoden der standardisierten Datenerhebung und Datenauswertung* (11., überarb. Aufl.). Stuttgart: Lucius & Lucius.
- Kuckartz, U. (2009a). Inhaltsanalyse. In B. Westle (Hrsg.), *Methoden der Politikwissenschaft* (S. 334-344). Baden-Baden: Nomos.
- Kuckartz, U. (2009b). Methodenkombination. In B. Westle (Hrsg.), *Methoden der Politikwissenschaft* (S. 352-362). Baden-Baden: Nomos.
- Kuckartz, U. (2010a). *Einführung in die computergestützte Analyse qualitativer Daten* (3., aktual. Aufl.). Wiesbaden: VS Verlag.
- Kuckartz, U. (2010b). Typenbildung. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 553-568). Wiesbaden: VS Verlag.
- Kuckartz, U. (2012). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz Juventa.
- Kuckartz, U. & Busch, J. (2012). Mixed Methods Evaluationen. In U. Kuckartz & S. Rädiker (Hrsg.), *Erziehungswissenschaftliche Evaluationspraxis. Beispiele – Konzepte – Methoden* (S. 145-158). Weinheim: Beltz Juventa.
- Kuckartz, U., Dresing, T., Rädiker, S. & Stefer, C. (2008). *Qualitative Evaluation. Der Einstieg in die Praxis* (2., aktual. Aufl.). Wiesbaden: VS Verlag.
- Kuckartz, U., Ebert, T., Rädiker, S. & Stefer, C. (2009). *Evaluation online. Internetgestützte Befragung in der Praxis*. Wiesbaden: VS Verlag.
- Kuckartz, U. & Rädiker, S. (2010a). Computergestützte Analyse (CAQDAS). In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 734-750). Wiesbaden: VS Verlag.
- Kuckartz, U. & Rädiker, S. (2010b). Computergestützte Verfahren. In K. Bock (Hrsg.), *Handbuch qualitative Methoden in der Sozialen Arbeit* (S. 353-362). Opladen u.a.: Budrich.

- Kuckartz, U. & Rädiker, S. (2012). *Erziehungswissenschaftliche Evaluationspraxis. Beispiele – Konzepte – Methoden*. Weinheim: Beltz Juventa.
- Kuper, H. (2005). *Evaluation im Bildungssystem*. Stuttgart: Kohlhammer.
- Kuper, H. (2008). Evaluationswissen als Steuerungsmedium. In S. Hartz & J. Schrader (Hrsg.), *Steuerung und Organisation in der Weiterbildung* (S. 311-325). Bad Heilbrunn: Klinkhardt.
- Kuwan, H., Bilger, F., Gnahn, D. & Seidel, S. (2006). Berichtssystem Weiterbildung IX. Integrierter Gesamtbericht zur Weiterbildungssituation in Deutschland. Bonn, Berlin: Bundesministerium für Bildung und Forschung. Verfügbar unter: [http://www.bmbf.de/pub/berichtssystem\\_weiterbildung\\_neun.pdf](http://www.bmbf.de/pub/berichtssystem_weiterbildung_neun.pdf) [13.06.12].
- Lüders, C. (2006). Qualitative Evaluationsforschung – Was heißt hier Forschung? In U. Flick (Hrsg.), *Qualitative Evaluationsforschung. Konzepte, Methoden, Umsetzungen*. (S. 33-62). Reinbek bei Hamburg: Rowohlt.
- Mark, M. M. & Henry, G. T. (2004). The Mechanisms and Outcomes of Evaluation Influence. *Evaluation*, 10(1), 35-57.
- Markowitsch, J. (2009). *Präsentationsunterlagen im Rahmen der Lehrveranstaltung „Einführung in Bildungscontrolling“, Sommersemester 2009*. Donau-Universität Krems.
- Marotzki, W. (2003). Leitfadeninterview. In R. Bohnsack, W. Marotzki & M. Meuser (Hrsg.), *Hauptbegriffe Qualitativer Sozialforschung. Ein Wörterbuch*. Opladen: Leske+Budrich.
- Mayring, P. (2002). *Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken* (5., überarb. und neu ausgestattete Aufl.). Weinheim u.a.: Beltz.
- Mayring, P. (2007). Generalisierung in qualitativer Forschung. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 8(3). Verfügbar unter: <http://www.qualitative-research.net/index.php/fqs/article/download/291/640> [18.04.12].
- Meier, R. (2009). *Praxis Bildungscontrolling. Was Sie wirklich tun können, um Ihre Aus- und Weiterbildung qualitätsbewusst zu steuern*. Offenbach: Gabal.
- Merriam, S. B. (1988). *Case Study Research in Education. A Qualitative Approach*. San Francisco: Jossey-Bass.
- Mertens, D. M. (1998). *Research methods in education and psychology. Integrating diversity with quantitative & qualitative approaches*. Thousand Oaks, CA u.a.: Sage.
- Mertens, D. M. (2008). *Transformative Research and Evaluation*. New York: Guilford Press.
- Meyer, J. (2006). Nach der Testierung ist vor der (Re)Testierung – Erfahrungen bei der Einführung des LQW-Qualitätsentwicklungssystems bei der NILES AW gGmbH. In R. Zech (Hrsg.), *Handbuch Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW). Grundlegung, Anwendung, Wirkung* (S. 221-229). Bielefeld: W. Bertelsmann.
- Meyer, W. (2007). Datenerhebung: Befragungen – Beobachtungen – Nicht-Reaktive Verfahren. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 223-277). Münster u.a.: Waxmann.
- Nagel, B. & Tiedtke, E. (2007). Das Rechtssystem in der Weiterbildung. In P. Krug & E. Nuissl (Hrsg.), *Praxishandbuch Weiterbildungsrecht*. Neuwied: Luchterhand.
- Nielsen, S. B., Lemire, S. & Skov, M. (2011). Measuring Evaluation Capacity – Results and Implications of a Danish Study. *American Journal of Evaluation*, 32(3), 324-344.
- OECD. (2003). *Definition und Auswahl von Schlüsselkompetenzen. Zusammenfassung des Abschlussberichts des Projekts „Definition und Selection of Competencies (DeSeCo)“*. Verfügbar unter: <http://www.oecd.org/dataoecd/36/56/35693281.pdf> [04.01.12].
- Owen, J. M. & Rogers, P. J. (2006). *Program evaluation. Forms and approaches*. London u.a.: Sage.
- Patton, M. Q. (1978). *Utilization-Focused Evaluation*. Beverly Hills, CA u.a.: Sage.
- Patton, M. Q. (1997). *Utilization-Focused Evaluation* (3rd ed.). Thousand Oaks, CA u.a.: Sage.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation* (4th ed.). Los Angeles u.a.: Sage.
- Perrin, B. (2005). How can information about the competencies required for evaluation be useful? *The Canadian Journal of Program Evaluation*, 20(2), 169-188.
- Philipps, J. & Schirmer, F. (2005). *Return on Investment in der Personalentwicklung. Der 5-Stufen-Evaluationsprozess*. Berlin: Springer.
- Picciotto, R. (2009). *Towards harmonized competencies for evaluation in Europe*.
- Porst, R. (2009). *Fragebogen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag.
- Preskill, H. & Torres, R. T. (2001). The Readiness for Organisational Learning and Evaluation Instrument. In D. Russ-Eft & H. Preskill (2008) (Hrsg.), *Evaluation in organisations: A systematic approach to enhancing learning, performance and change* (S. 428-433). Cambridge, MA: Perseus.
- Rädiker, S. (2006). Die Wirkungen der Lernerorientierten Qualitätstestierung in der Praxis der Weiterbildungsorganisationen. In R. Zech (Hrsg.), *Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW)* (S. 201-220). *Grundlegung. Anwendung. Wirkung*. Bielefeld: W. Bertelsmann.
- Rädiker, S. (2009). Wie kann pädagogische Qualität gemessen werden. Evaluation von Bildungsprozessen. In C. Dehn (Hrsg.), *Pädagogische Qualität. Einflussfaktoren und Wirkmechanismen*. (S. 132-153). Hannover: Expressum.
- Rädiker, S. (2010a). *Die Inter-Coder-Reliabilität. Überlegungen zu Codierer-Übereinstimmungen bei der computergestützten Analyse qualitativer Daten*. Paper vorgestellt auf der 12. CAQD-Tagung, Mar-

- burg. Verfügbar unter: [http://www.caqd.de/dokumente/vortraege-2010/CAQD\\_2010\\_Inter-Coder\\_Raediker.pdf](http://www.caqd.de/dokumente/vortraege-2010/CAQD_2010_Inter-Coder_Raediker.pdf) [12.04.12].
- Rädiker, S. (2010b). *Lohnt sich die (Re-)Testierung? Auswirkungen der Qualitätsentwicklung nach LQW bei Ersttestierern und Retestierern*. Verfügbar unter: [http://www.artset.de/cms/fileadmin/user\\_upload/2010\\_10\\_24\\_Wirkungsanalyse\\_LQW3.pdf](http://www.artset.de/cms/fileadmin/user_upload/2010_10_24_Wirkungsanalyse_LQW3.pdf) [04.04.12].
- Rädiker, S. (2010c). Professionelle Erfolgskontrolle: Wie beschreiben Weiterbildungsorganisationen ihre Evaluation der Bildungsprozesse? In R. Zech, C. Dehn, K. Tödt, S. Rädiker, M. Mrugalla & J. Schunter (Hrsg.), *Organisationen in der Weiterbildung. Selbstbeschreibungen und Fremdbeschreibungen* (S. 132-168). Wiesbaden: VS Verlag.
- Rädiker, S. (2011). Die Lernerorientierte Qualitätstestierung auf dem Prüfstand. Ergebnisse einer Vorher-Nachher-Befragung. *MAGAZIN erwachsenenbildung.at. Das Fachmedium für Forschung, Praxis und Diskurs*, 2011(12), 05-01 - 05-09. Verfügbar unter: <http://www.erwachsenenbildung.at/magazin/11-12/meb11-12.pdf>
- Rädiker, S. (2012). Die Evaluation von Weiterbildungsprozessen. Eine Analyse der Herausforderungen in LQW-testierten Organisationen. In U. Kuckartz & S. Rädiker (Hrsg.), *Erziehungswissenschaftliche Evaluationspraxis. Beispiele – Konzepte – Methoden* (S. 35-50). Weinheim: Beltz Juventa.
- Rädiker, S. & Niklesz, A. (2012). Kompetenzen von Evaluierenden. In U. Kuckartz & S. Rädiker (Hrsg.), *Erziehungswissenschaftliche Evaluationspraxis. Beispiele – Konzepte – Methoden* (S. 200-220). Weinheim: Beltz Juventa.
- Rädiker, S. & Stefer, C. (2007). Qualitative Evaluation – Versuch einer Abkürzungsstrategie. In U. Kuckartz, H. Grunenberg & T. Dresing (Hrsg.), *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis* (2., überarb. und erw. Auflage, S. 78-92). Wiesbaden: VS Verlag.
- Reischmann, J. (2004). Warum nicht mal den Nutzen messen? Evaluation von Bildungsarbeit. *DIE-Zeitschrift für Erwachsenenbildung*, 2004(3), 42-48.
- Reischmann, J. (2006). *Weiterbildungs-Evaluation. Lernerfolge messbar machen* (2. Aufl.). Augsburg: Ziel.
- Rindermann, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2(2), 233-256.
- Rossi, P. H., Freeman, H. E. & Lipsey, M. W. (2006). *Evaluation. A Systematic Approach* (7th ed.). London u.a.: Sage.
- Schmidt, U. (2002). Methodenkompetenz in der Evaluation. *Zeitschrift für Evaluation*, 1(1), 197-202.
- Schnell, R., Hill, P. B. & Esser, E. (2008). *Methoden der empirischen Sozialforschung* (8. Aufl.). München: Oldenbourg.
- Schöni, W. (2009). *Handbuch Bildungscontrolling. Steuerung von Bildungsprozessen in Unternehmen und Bildungsinstitutionen* (2., aktual. und erw. Aufl.). Zürich u.a.: Rüegger.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA u.a.: Sage.
- Scriven, M. (1996). Types of Evaluation and Types of Evaluator. *Evaluation Practice*, 17(2), 151-161.
- Scriven, M. (2007). *Key Evaluation Checklist. Intended for use in designing and evaluating programs, plans, and policies; writing evaluation reports on them; assessing their evaluability; and evaluating evaluations of them*. Verfügbar unter: [http://www.wmich.edu/evalctr/archive\\_checklists/kec\\_feb07.pdf](http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf) [18.04.12].
- Seale, C. (1999). *The quality of qualitative research*. Thousand Oaks, CA u.a.: Sage.
- Seeber, S. (2000). Stand und Perspektiven von Bildungscontrolling. In S. Seeber, E. M. Krekel & J. van Buer (Hrsg.), *Bildungscontrolling – Ansätze und kritische Diskussionen zur Effizienzsteigerung von Bildungsarbeit*. (S. 19-50). Frankfurt am Main u.a.: Lang.
- Simonson, J. & Pötschke, M. (2006). Akzeptanz internetgestützter Evaluationen an Universitäten. *Zeitschrift für Evaluation*, 5(2), 227-248.
- Stake, R. E. (1973). *Program Evaluation, Particularly Responsive Evaluation*. Paper vorgestellt auf der Tagung „New Trends in Evaluation“, Institute of Education at Göteborg University. Verfügbar unter: <https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/05%20Program%20Evaluation,%20Particularly%20Responsive%20Evaluation.pdf> [28.03.12].
- Stake, R. E. (2004). *Standards-Based and Responsive Evaluation*. Thousand Oaks, CA u.a.: Sage.
- Stefer, C. (in Vorbereitung). *Die Gegenstandsangemessenheit empirischer Datenerhebungsmethoden im Kontext von Lehrevaluationen an Hochschulen*. Doktorarbeit, Philipps-Universität Marburg.
- Steinke, I. (2007). Qualitätssicherung in der qualitativen Forschung. In U. Kuckartz, H. Grunenberg & T. Dresing (Hrsg.), *Qualitative Datenanalyse: computergestützt. Methodische Hintergründe und Beispiele aus der Forschungspraxis* (2. Aufl., S. 176-187). Wiesbaden: VS Verlag.
- Stevahn, L., King, J. A., Ghore, G. & Minnema, J. (2005). Establishing Essential Competencies for Program Evaluators. *American Journal of Evaluation*, 26(1), 43-59.
- Stockbauer, U. (2000). Was macht Evaluationen nützlich? Überblick zum Forschungsstand – Ergebnisse von Fallstudien. In H. Müller-Kohlenberg & K. Münstermann (Hrsg.), *Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen* (S. 121-128). Opladen: Leske+Budrich.
- Stockmann, R. (2004). *Evaluationsforschung. Grundlagen und ausgewählte Forschungsfelder* (2., überarb. und aktual. Aufl.). Opladen: Leske+Budrich.



- Stockmann, R. (2006). *Evaluation und Qualitätsentwicklung. Eine Grundlage für wirkungsorientiertes Qualitätsmanagement*. Münster u.a.: Waxmann.
- Stockmann, R. (2007). *Handbuch zur Evaluation. Eine praktische Handlungsanleitung*. Münster u.a.: Waxmann.
- Stockmann, R. (2010a). Einleitung. In R. Stockmann & W. Meyer, *Evaluation. Eine Einführung* (S. 9-14). Opladen u.a.: Barbara Budrich.
- Stockmann, R. (2010b). Rolle der Evaluation in der Gesellschaft. In R. Stockmann & W. Meyer, *Evaluation. Eine Einführung* (S. 15-54). Opladen u.a.: Barbara Budrich.
- Stockmann, R. (2010c). Wissenschaftsbasierte Evaluation. In R. Stockmann & W. Meyer, *Evaluation. Eine Einführung* (S. 55-100). Opladen u.a.: Barbara Budrich.
- Stockmann, R. & Meyer, W. (2010). *Evaluation. Eine Einführung*. Opladen u.a.: Barbara Budrich.
- Stufflebeam, D. L. (1969). Evaluation as enlightenment for decision-making. In H. B. Walcott (Hrsg.), *Improving educational assessment and an inventory of measures of affective behavior* (S. 41-73). Washington, DC: Association for Supervision and Curriculum Development and National Education Association (dt. in Wulf, 1972).
- Stufflebeam, D. L. (1971). *Educational evaluation and decision making*. Itasca, IL: Peacock.
- Stufflebeam, D. L. (1972). Evaluation als Entscheidungshilfe. In C. Wulf (Hrsg.), *Evaluation. Texte. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 113-145). München: R. Piper & Co.
- Stufflebeam, D. L. (2003). *THE CIPP MODEL FOR EVALUATION. An update. A review of the model's development. A checklist to guide implementation*. Paper vorgestellt auf der Annual Conference of the Oregon Program Evaluators Network (OPEN) Portland, Oregon. Verfügbar unter: <http://goeroendeso.files.wordpress.com/2009/01/cipp-modeloregon10-031.pdf> [12.04.12].
- Stufflebeam, D. L. (2007). *CIPP EVALUATION MODEL CHECKLIST. A tool for applying the CIPP Model to assess long-term enterprises. Intended for use by evaluators and evaluation clients/stakeholders [2nd ed.]*. Verfügbar unter: [http://www.wmich.edu/evalctr/archive\\_checklists/cippchecklist\\_mar07.pdf](http://www.wmich.edu/evalctr/archive_checklists/cippchecklist_mar07.pdf) [17.03.12].
- Stufflebeam, D. L., Madaus, G. F. & Kellaghan, T. (2000). *Evaluation models. Viewpoints on educational and human services evaluation* (2nd ed.). Boston u.a.: Kluwer.
- Stufflebeam, D. L. & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. San Francisco: Jossey-Bass.
- Suchman, E. A. (1967). *Evaluative Research. Principles and Practice in Public Service & Social Action Programs*. New York: Russell Sage Foundation.
- Taut, S. & Brauns, D. (2003). Resistance to Evaluation. A Psychological Perspective. *Evaluation*, 9(3), 247-264.
- Taylor, P. & Beniast, J. (2003). *Training in agroforestry. A toolkit for trainers*. Verfügbar unter: <http://www.cglrc.cgiar.org/icraf/toolkit/toolkit.pdf> [22.03.12].
- Tietgens, H. (1971). Methoden > Arbeitskontrollen. In Pädagogische Arbeitsstelle des Deutschen Volkshochschul-Verbandes (Hrsg.), *Die Volkshochschule. Handbuch für die Praxis der Leiter und Mitarbeiter. Lose-Blatt-Sammlung. 1968-1991*. (Blatt 73.000). Frankfurt am Main.
- Tietgens, H. (1984). Methoden > Arbeitskontrollen > Formen der Evaluation. In Pädagogische Arbeitsstelle des Deutschen Volkshochschul-Verbandes (Hrsg.), *Die Volkshochschule. Handbuch für die Praxis der Leiter und Mitarbeiter. Lose-Blatt-Sammlung. 1968-1991*. (Blatt 73.101). Frankfurt am Main.
- Tippelt, R. & von Hippel, A. (2010). *Handbuch Erwachsenenbildung/Weiterbildung* (4. durchgesehene Aufl.). Wiesbaden: VS Verlag.
- UNEG. (2008). *Core Competencies for Evaluators of the UN System*. Verfügbar unter: [http://www.unevaluation.org/documentdownload?doc\\_id=81&file\\_id=559](http://www.unevaluation.org/documentdownload?doc_id=81&file_id=559) [04.04.12].
- Van den Berghe, W. (1998). Application of ISO 9000 standards to education and training. *Vocational training European journal*, 3(15), 20-28.
- Vedung, E. (1999). *Evaluation im öffentlichen Sektor*. Wien: Böhlau.
- Veltjens, B. (2006). *Qualitätsmodelle im Überblick*. Verfügbar unter: [http://www.die-bonn.de/esprid/dokumente/doc-2006/veltjens06\\_01.pdf](http://www.die-bonn.de/esprid/dokumente/doc-2006/veltjens06_01.pdf) [04.04.12].
- Veltjens, B. (2010). Qualitätsmanagement. In R. Arnold, S. Nolda & E. Nuissl (Hrsg.), *Wörterbuch Erwachsenenbildung* (2., überarb. Aufl., S. 252-255). Bad Heilbrunn: Klinkhardt.
- von Ameln, F. & Zech, R. (2011). Musterwechsel in Organisationen. Latente Organisationsregeln als Schlüsselfaktoren gelingenden Change Managements. *OrganisationsEntwicklung. Zeitschrift für Unternehmensentwicklung und Change Management*, 4(30), 49-55.
- von Wensierski, H.-J. (2003). Rekonstruktive Sozialpädagogik im intermediären Feld eines Wissenschaft-Praxis-Diskurses – am Beispiel Praxisforschung. In C. Schweppe (Hrsg.), *Qualitative Forschung in der Sozialpädagogik* (S. 67-90). Opladen: Leske+Budrich.
- W.K. Kellogg Foundation. (2004). *Logic Model Development Guide. Using Logic Models to Bring Together Planning, Evaluation, and Action. Updated January 2004*. Verfügbar unter: <http://www.wkcf.org/~media/36693510092544928C454B5778180D75/LogicModel.pdf> [28.03.12].
- Weber, S. M. (2012). Transformative Evaluation. In U. Kuckartz & S. Rädiker (Hrsg.), *Erziehungswissenschaftliche Evaluationspraxis. Beispiele – Konzepte – Methoden* (S. 123-144). Weinheim: Beltz Juventa.

- Weinberg, J. (2000). *Einführung in das Studium der Erwachsenenbildung* (überarb. Neuauflage). Bad Heilbrunn: Klinkhardt.
- Wesseler, M. (2010). Evaluation und Evaluationsforschung. In R. Tippelt & A. von Hippel (Hrsg.), *Handbuch Erwachsenenbildung/Weiterbildung* (4., durchgesehene Aufl., S. 1031-1048). Wiesbaden: VS Verlag.
- Widmer, T. (2000). Kontext, Inhalt und Funktion der „Standards für die Evaluation von Programmen“. In H. Müller-Kohlenberg & K. Münstermann (Hrsg.), *Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen* (S. 77-88). Opladen: Leske+Budrich.
- Will, H. & Blickhan, C. (1987). Evaluation als Intervention. In H. Will, A. Winteler & A. Krapp (Hrsg.), *Evaluation in der beruflichen Aus- und Weiterbildung*, (S. 43-59). Heidelberg: Sauer.
- Will, H., Winteler, A. & Krapp, A. (1987a). Von der Erfolgskontrolle zur Evaluation. In H. Will, A. Winteler & A. Krapp (Hrsg.), *Evaluation in der beruflichen Aus- und Weiterbildung* (S. 11-42). Heidelberg: Sauer.
- Will, H., Winteler, A. & Krapp, A. (Hrsg.). (1987b). *Evaluation in der beruflichen Aus- und Weiterbildung. Konzepte und Strategien*. Heidelberg: Sauer.
- Windham, D. M. (1988). *Improving the Efficiency of Educational Systems*. Verfügbar unter: [http://pdf.usaid.gov/pdf\\_docs/PNABA775.pdf](http://pdf.usaid.gov/pdf_docs/PNABA775.pdf) [18.04.12].
- Wittmann, W. W. (1985). *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Berlin u.a.: Springer.
- Witzel, A. (2000). Das problemzentrierte Interview [26 Absätze]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(1). Verfügbar unter: <http://www.qualitative-research.net/fqs-texte/1-00/1-00witzel-d.htm> [18.10.11].
- Wottawa, H. (1996). Wirksamkeit von Trainingsmaßnahmen. Evaluationsforschung bei Weiterbildungsveranstaltungen. *Grundlagen der Weiterbildung*, 7(1), 4-9.
- Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2., vollständig überarb. Aufl.). Bern u.a.: Huber.
- Wottawa, H. & Thierau, H. (2003). *Lehrbuch Evaluation* (3., korr. Aufl.). Bern u.a.: Huber.
- Wulf, C. (Hrsg.). (1972). *Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen*. München: R. Piper & Co.
- Zech, R. unter Mitarbeit von Angermüller, J., Ehse, C., Erhart, F., Harp, S., Kieselhorst, M., Meyer, J., et al. (2006). *Handbuch Lernerorientierte Qualitätstestierung in der Weiterbildung (LQW). Grundlegung, Anwendung, Wirkung*. Bielefeld: W. Bertelsmann.
- Zech, R. unter Mitarbeit von von Ameln, F., Brückner, F., Dehn, C., Erhart, F., Grüning, G. & Hagedorn, K. (2010). *Handbuch Management in der Weiterbildung*. Weinheim u.a.: Beltz.
- Zech, R. (2006). *Lernerorientierte Qualitätstestierung in der Weiterbildung. Leitfaden für die Praxis*. Hannover: Expressum.
- Zech, R. (2008). *Handbuch Qualität in der Weiterbildung*. Weinheim u.a.: Beltz.
- Zech, R. (2010). Organisationen in der Weiterbildung. Selbstbeschreibungen und Fremdbeschreibungen. In R. Zech, C. Dehn, K. Tödt, S. Rädiker, M. Mrugalla & J. Schunter (Hrsg.), *Organisationen in der Weiterbildung. Selbstbeschreibungen und Fremdbeschreibungen* (S. 14-71). Wiesbaden: VS Verlag.
- Zech, R. (Hrsg.). (2004). *Qualität durch Reflexivität. Lernerorientierte Qualitätsentwicklung in der Praxis*. Hannover: Expressum.
- Zech, R. & Braucks, D. (2004). Qualität durch Reflexivität. Lernerfolge, Entwicklungsbedarfe und Erfolgsfaktoren der Qualitätsentwicklung. In R. Zech (Hrsg.), *Qualität durch Reflexivität* (S. 11-38). Hannover: Expressum.
- Zech, R., Dehn, C., Tödt, K., Rädiker, S., Mrugalla, M. & Schunter, J. (2010). *Organisationen in der Weiterbildung. Selbstbeschreibungen und Fremdbeschreibungen*. Wiesbaden: VS Verlag.
- Zech, R. & Tödt, K. (2012). *Gelungenes Lernen – Qualität und Qualitätsmanagement in der Weiterbildung*. Technische Universität Kaiserslautern.
- Zürcher, R. (2010). Kompetenz – eine Annäherung in fünf Schritten. *MAGAZIN erwachsenenbildung.at. Das Fachmedium für Forschung, Praxis und Diskurs*, 2010(9), 04-01 - 04-12. Verfügbar unter: <http://erwachsenenbildung.at/magazin/10-09/meb10-9.pdf>
- Zuschlag, B. (1987). Widerstände gegen Evaluationsmaßnahmen. In H. Will, A. Winteler & A. Krapp (Hrsg.), *Evaluation in der beruflichen Aus- und Weiterbildung* (S. 75-88). Heidelberg: Sauer.

## Anhang

### A) Kategorienhandbuch für die Analyse der Selbstreporte

Die folgende Tabelle enthält die Definitionen der Hauptkategorien, die für die Analyse der Selbstreporte verwendet wurden. Die Definitionen sind teilweise noch um Ankerbeispiele ergänzt, die hier nicht abgedruckt sind.

01 Evaluationszweck	<p>In diese Kategorie fallen alle Aussagen, die das Ziel, den Zweck oder den Sinn der durchgeführten Evaluation/en in der Organisation betreffen. Es handelt sich also um Aussagen, welche die Frage beantworten „Warum evaluieren Sie?“</p> <p>Schlüsselwörter sind: Mit der Evaluation „sollen“ oder „wollen wir“ ... oder auch Evaluation „heißt für uns“ ... oder „Ziel ist es ...“</p> <p>Abgrenzung zur Kategorie „08c Nutzung/Nutzen“: Erst, wenn es über einen Informationsgewinn hinausgeht, wird Kategorie 08c zugeordnet.</p> <p>Aufgrund der LQW-Mindestanforderungen kommen in diese Kategorie auch Begründungen für die Anwendung von Evaluationsverfahren, sofern die Begründungen etwas Fundiertes darüber aussagen, wozu die Verfahren eingesetzt werden und somit die Frage beantworten, wozu die Evaluation dient.</p>
02 Gegenstände/Inhalt	<p>In diese Kategorie fallen alle Aussagen, die die Frage beantworten "Was evaluieren Sie?". Es kann sich handeln um</p> <ul style="list-style-type: none"><li>a) die Bereiche, die in einem Fragebogen abgefragt werden,</li><li>b) eine Auflistung der Gesprächsthemen von Interviews mit Teilnehmenden,</li><li>c) eine allgemeinere Aufzählung oder</li><li>d) Fragestellungen, die eine Organisation mit der Evaluation beantworten möchte.</li></ul> <p>Es wird nur das codiert, was intendiert abgefragt oder erhoben wird.</p> <p>Es wird für jede Datenerhebungsmethode codiert, welche Gegenstände vorkommen – es kann also zu Doppelcodierungen innerhalb eines Dokumentes kommen.</p>
03 Instrumente/Datenquellen	<p>In diese Kategorie fallen alle Aussagen über a) Instrumente und Methoden, mit denen Daten für die Evaluation gewonnen werden. Manchmal werden auch nur b) Datenquellen benannt. Dabei werden alle von einer Organisation genannten Instrumente/Datenquellen codiert, auch wenn aus evaluationstheoretischer Sicht diese evtl. nicht als klassische Instrumente zu sehen sind.</p>
04 Design	<p>In diese Kategorie fallen alle Angaben über</p> <ul style="list-style-type: none"><li>a) den Zeitpunkt der Evaluation bzw. über den Zeitpunkt des Einsatzes von Evaluationsinstrumenten,</li><li>b) Rhythmus, also wie häufig und ggf. in welchem Turnus Datenerhebungen stattfinden,</li><li>c) Umfang der Evaluationen, z.B. die Information, ob ein Kurs, mehrere Kurse oder alle Kurse evaluiert werden. Umfang meint NICHT, wie viele Seiten ein Evaluationsbogen hat.</li></ul> <p>Auch d) Begründungen des Designs, v.a. des Umfangs, sind hier zu codieren</p>
05 Datenerhebung	<p>In diese Kategorie fallen alle Aussagen über die Erhebung von Daten, wobei es nicht um die Instrumente aus Kategorie 03 geht, sondern um das <u>Procedere</u>, wie die Datenerhebung abläuft.</p>

06 Datenauswertung und -analyse	<p>In diese Kategorie fallen alle Aussagen, welche die Schlüsselwörter "Auswertung", "auswerten" oder "Analyse", "analysieren" enthalten, unabhängig davon, ob klar wird, wie ausgewertet wird.</p> <p>Abgrenzung: Die reine Übertragung von Daten auf ein Formblatt wird noch nicht als Auswertung, sondern noch als "06 Datenerhebung" betrachtet.</p> <p>Abgrenzung: Ein Auswertungsgespräch im Sinne von Reflektion ist keine Auswertung, es sei denn, im Gespräch wird ausgewertet.</p> <p>Datenauswertung vs. Ergebnisauswertung: Daten werden ausgewertet, dies führt zu Ergebnissen, diese können wiederum ausgewertet werden.</p>
07 Datenbewertung	<p>In diese Kategorie fallen alle Aussagen über die Bewertung der erhobenen Daten.</p> <p>Aussagen über Bewertungen der Teilnehmenden (die nicht durch die Einrichtung bewertet wird) werden nicht codiert.</p> <p>Abgrenzung: Bewertungskriterien/Indikatoren werden unter "02 Gegenstände ..." codiert.</p>
08a Umgang mit Ergebnissen > Dokumentation	<p>In diese Kategorie fallen alle Aussagen, die darüber informieren, wie die Organisationen ihre Ergebnisse oder Erhebungen dokumentieren. Dazu gehören vor allem Dokumente, die physisch und/oder virtuell vorliegen.</p>
08b Umgang mit Ergebnissen > Kommunikation/Information	<p>In diese Kategorie fallen alle Aussagen, die darüber Aufschluss geben, wie die Ergebnisse der Evaluationen (in welcher Form sie auch vorliegen mögen) an</p> <ul style="list-style-type: none"> <li>a) Lehrende</li> <li>b) Teilnehmende</li> <li>c) Mitarbeitende und Leitung oder</li> <li>d) Auftraggebende und Träger sowie</li> <li>e) die Öffentlichkeit</li> </ul> <p>kommuniziert werden bzw. wie mit diesen Personen(gruppen) über die Ergebnisse diskutiert wird.</p>
08c Umgang mit Ergebnissen > Nutzung/Konsequenzen	<p>In diese Kategorie fallen alle Aussagen, die darüber Aufschluss geben, wie die Ergebnisse der Evaluationen konkret verwendet werden (Nutzung – nicht Nutzen).</p> <p>Der Nutzen, im Sinne von einer Absichtserklärung oder eines reinen Informationsgewinns wird bei 01 Evaluationszweck codiert (dieser ist in der Regel durch Wörter wie "soll", "dient" etc. gekennzeichnet).</p> <p>In diese Kategorie gehören ebenso Aussagen, die über den Umgang mit Ergebnissen im Sinne von Konsequenzen informieren.</p>
Problemfelder/Entwicklungsbedarfe	<p>In diese Kategorie fallen alle Aussagen, in denen ein Problem im Rahmen der Evaluation, Entwicklungsbedarfe oder konkrete Entwicklungs-/Veränderungspläne zum Ausdruck kommen.</p> <p>Hierunter fallen sowohl a) bewältigte Probleme (Aussagen im Perfekt/Imperfekt) als auch b) aktuell bestehende.</p>

## B) Leitfaden für die Interviews mit Evaluationsverantwortlichen

Nachfragemöglichkeiten sind in eckige Klammern gesetzt.

<b>Einleitung</b>
<p>Guten Tag,</p> <p>vielen Dank noch einmal an dieser Stelle, dass Sie sich für das Interview bereit erklärt haben und auch schon den Online-Fragebogen ausgefüllt haben.</p> <p>Zu Beginn möchte ich noch einmal kurz das Ziel unserer Studie benennen: Uns interessiert, wie in Ihrer Organisation die Bildungsprozesse evaluiert werden. Uns interessiert dabei vor allem, welche Herausforderungen Sie dabei zu bewältigen haben, welche Entwicklungspotenziale Sie sehen und wie die Evaluationsergebnisse verwendet werden.</p> <p>Das Interview wird voraussichtlich 30 Minuten dauern.</p> <p>Wie bereits geschrieben, zeichnen wir die Interviews auf, damit für die spätere Analyse keine wichtigen Informationen verloren gehen. Natürlich werden alle Ihre Angaben vertraulich behandelt.</p> <p>Das Interview teilt sich in drei Themenblöcke:</p> <ol style="list-style-type: none"> <li>1. Herausforderungen und Entwicklungspotenziale</li> <li>2. Nutzung der Evaluation und</li> <li>3. Kompetenzen und Know-how</li> </ol> <p>Dann möchte ich jetzt mit dem ersten Themenblock beginnen.</p>
<b>Herausforderungen und Entwicklungspotenziale</b>
<ol style="list-style-type: none"> <li>1. Wenn Sie mal an den Qualitätsbereich 5 von LQW zur Evaluation der Bildungsprozesse denken. Wo sehen Sie in Ihrer Organisation Herausforderungen, Unsicherheiten oder Schwierigkeiten in Bezug auf die Evaluation der Bildungsprozesse? [Gibt es noch etwas Weiteres, wo Sie Herausforderungen oder Schwierigkeiten sehen? (ggf. 2x nachfragen)]</li> <li>2. Ich nenne Ihnen nun einige Bereiche, in denen Probleme bei der Evaluation auftreten können. Ich bitte Sie, jeweils kurz zu sagen, wie es in dem Bereich in Ihrer Organisation aussieht: <ol style="list-style-type: none"> <li>a) Akzeptanz der Evaluation durch Lehrende, Teilnehmende oder Mitarbeitende</li> <li>b) Erhebungsinstrumente und -verfahren</li> <li>c) Erhebung von Daten</li> <li>d) Auswertung von Daten</li> <li>e) Arbeitsaufwand</li> </ol> </li> <li>3. Welche von den Unsicherheiten oder Schwierigkeiten, die Sie genannt haben, stellt für Sie das größte Problem dar? [Und warum?]</li> <li>4. Inwieweit ist in Ihrer Organisation beabsichtigt, die derzeitige Evaluationspraxis zukünftig zu verändern? [Sind Qualitätsentwicklungsziele oder strategische Entwicklungsziele im Rahmen von LQW definiert, in denen die Evaluation der Bildungsprozesse eine Rolle spielt?]</li> <li>5. Inwieweit sehen Sie in Ihrer Organisation Unterstützungsbedarf bei der Evaluation von Bildungsprozessen, z.B. durch externe Personen?</li> </ol>
<b>Nutzung der Evaluation</b>
<ol style="list-style-type: none"> <li>6. Mit welchem Ziel werden in Ihrer Organisation eigentlich die Bildungsprozesse evaluiert?</li> <li>7. Welche Ergebnisse erhält Ihre Organisation durch die Evaluation der Bildungsprozesse? In welcher Form liegen die Evaluationsergebnisse vor? [Wie kommen Sie von den Daten zu Konsequenzen? Welche Rolle spielen Bewertungskriterien und Sollwerte? Wie kommen Sie von den Daten zur Umsetzung?]</li> <li>8. Wie werden in Ihrer Organisation die Evaluationsergebnisse genutzt?</li> <li>9. Wer wird über die Ergebnisse der Evaluation informiert? [Teilnehmende, Dozenten, Mitarbeiter/innen, Leitung, Träger, Öffentlichkeit, Wissenschaft; Warum werden diese Personen informiert? Wie? (Internet, Zeitung, Bericht etc.?)]</li> </ol>

10. Sie haben ja jetzt verschiedene Aspekte genannt, wie Sie die Evaluation und ihre Ergebnisse nutzen. Wenn Sie so insgesamt auf die Evaluation der Bildungsprozesse schauen: Wie beurteilen Sie den Nutzen der Evaluation insgesamt? [Wie beurteilen Sie den Aufwand für die Evaluation im Verhältnis zum Nutzen?]
11. Was meinen Sie: Wo müsste man ansetzen, wenn man den Nutzen der Evaluation für Ihre Organisation erhöhen wollte?
<b>Kompetenzen und Know-how</b>
12. Welche Kompetenzen braucht man aus Ihrer Sicht, um Bildungsprozesse effizient zu evaluieren?
13. Welche dieser Kompetenzen sind in Ihrer Organisation vorhanden? [Wie haben Sie diese Kompetenzen erlangt?]
14. Mal angenommen, Sie hätten Zeit und Geld, ein Seminar zum Thema „Evaluation von Bildungsprozessen“ zu besuchen. Was würden Sie oder eine Person aus Ihrer Organisation dort lernen wollen? Was wäre für Sie bzw. für Ihre Organisation hilfreich?
15. Was verstehen Sie persönlich eigentlich unter Evaluation? Wie würden Sie also Evaluation definieren?
16. Wenn ich andere Personen in Ihrer Organisation fragen würde, was sie unter Evaluation verstehen: Würden diese Personen ähnlich antworten?
<b>Abschluss</b>
17. Gibt es noch etwas Wichtiges aus Ihrer Sicht zu sagen?

### C) Anschreiben an die Interviewpersonen

Sehr geehrte/r Frau/Herr XY,

ich bin wissenschaftlicher Mitarbeiter der Marburger Arbeitsgruppe für Methoden und Evaluation an der Uni Marburg. Ich schreibe derzeit meine Dissertation und arbeite hierfür gemeinsam mit der Diplomandin Julia Busch an dem Thema „Evaluation von Bildungsprozessen in LQW-testierten Organisationen“.

Uns interessiert, wie Weiterbildungsorganisationen ihre Bildungsprozesse evaluieren (QB 5 von LQW), welche Herausforderungen sie dabei bewältigen müssen und wie sie die Ergebnisse der Evaluation nutzen.

Wir bitten Sie herzlich um Ihre Unterstützung: Wir möchten gerne mit der Person in Ihrer Organisation, die mit der Evaluation vertraut ist, ein Telefoninterview führen. Das Interview dauert ca. 30 Minuten. Zur Studie gehört zudem ein kurzer Online-Fragebogen, der 5-10 Minuten Zeit beansprucht.

Würden Sie oder eine andere Person aus Ihrer Organisation für ein Interview zur Verfügung stehen? Wenn ja: Wann wäre ein geeigneter Telefontermin? Unser Wunsch ist es, die Interviews in den nächsten beiden Wochen abzuschließen.

Natürlich werden alle Ihre Angaben anonym und vertraulich behandelt.

Wir würden uns sehr über eine Zusage von Ihnen freuen.

Vielen Dank im Voraus für Ihre Mühe

Stefan Rädiker

## D) Kategorienhandbuch für die Analyse der Interviews

Die folgende Tabelle enthält die Definitionen der Hauptkategorien, die im Rahmen der vorliegenden Arbeit ausgewertet wurden (die Kategorien 05 bis 08 wurden in einem anderen Forschungsprojekt berücksichtigt und sind daher nicht in der Tabelle enthalten). Die Definitionen sind teilweise noch um Ankerbeispiele ergänzt, die hier nicht abgedruckt sind. Für die erste Kategorie sind zusätzlich die Definitionen der Subkategorien aufgeführt, da in die erste Kategorie die meisten Aussagen fallen.

01 Problemfelder der Evaluation	In diese Kategorie fallen alle Nennungen mit direktem oder indirektem Bezug zu Problemfeldern, die bei der Evaluation auftauchen können. Dies umfasst nicht nur a) die Aussagen zu konkreten Problemen, sondern auch b) die Aussagen, dass keine Probleme in bestimmten Bereichen (mehr) existieren.
01.1 Instrumente, Verfahren	Probleme, die sich auf die eingesetzten Evaluationsinstrumente und Evaluationsverfahren (=LQW-Terminologie) beziehen. Abgrenzung: Die Datenerhebung gehört nicht zum Evaluationsverfahren.
01.2 Datenerhebung	Probleme, welche die Datenerhebung im engeren Sinn betreffen: Verteilung und Ausfüllen von Fragebögen, Rückläufe etc.
01.3 Akzeptanz, Belastung	Aussagen, die a) Akzeptanzprobleme erkennen lassen, z.B. Widerstände, Vorbehalte oder b) auf Belastung hinweisen, z.B. zu viele, zu häufige, zu schwere Erhebungen. Der Code wird nur vergeben, wenn die Befragten es auch als Problem sehen. Der Code wird nicht vergeben, wenn gesagt wird, dass es Personen oder Personengruppen gebe, bei denen keine Akzeptanz da ist, das aber insgesamt gesehen kein Problem darstelle.
01.4 Datenauswertung und -bewertung	Probleme, welche die Auswertung (Softwareeinsatz, Aggregation etc.) oder die Bewertung von Daten betreffen.
01.5 Aufwand, Ressourcen	Probleme bezüglich Arbeitsbelastung, Effizienz, Effektivität, Aufwand-Nutzen-Relation, Aufwand allgemein, Belastung
01.6 Nutzen, Erkenntnisgewinn, Aussagekraft	Probleme, die den Nutzen der Evaluationspraxis betreffen, insbesondere Hinweise auf a) zu wenige, zu ungenaue oder zu wenig hilfreiche Informationen durch die Evaluation, b) aussagelose Daten, aber auch Probleme, die die Dimension c) Bewertung des Nutzens ansprechen.
01.7 Nutzung, Umsetzung, Ergebnisverwendung	Probleme zur Nutzung im Sinne von Umsetzung und Ergebnisverwendung.
01.8 Sonstiges	Probleme, die sich keiner vorherigen Kategorie zuordnen lassen und nicht häufiger als 4 Mal vorkommen. Nennungen, die mehr als einmal vorkommen, bekommen eine eigene Unterkategorie.
01.9 Bereiche ohne Probleme (mehr)	In diese Kategorie fallen sowohl Aussagen darüber, dass a) Probleme in bestimmten Bereichen nicht vorhanden sind, als auch, dass b) Probleme bewältigt wurden und nicht mehr existieren.

02 Beabsichtigte Veränderungen i.d. Evaluationspraxis	Aussagen über geplante Veränderungen der Evaluationspraxis, z.B. konkrete Qualitätsentwicklungsziele
03 Unterstützungsbedarf	Aussagen über Unterstützungsbedarf bei der Durchführung der Evaluation.
04 Ziele/Zweck der Evaluation	Aussagen zu Zielen und Zwecken der durchgeführten Evaluation. Als Haupt-Subkategorien wurden zu Vergleichszwecken die Kategorien der Analyse der Selbstreporte in der dortigen Reihenfolge übernommen.
09 Nutzenerhöhung	Aussagen darüber, wodurch der Nutzen der Evaluation in der Organisation erhöht werden kann.
10a Benötigte Kompetenzen für die Evaluation	Aussagen darüber, welche Kompetenzen benötigt werden, um Bildungsprozesse erfolgreich zu evaluieren. Die Subkategorien entsprechen den vier DeGEval Kompetenzfeldern und ihren jeweiligen Beispielen.
10b vorhandene Kompetenzen	Aussagen darüber, welche Kompetenzen in der Organisation bezüglich Evaluation vorhanden sind – insbesondere im Vergleich zu den Kompetenzen, die von den Befragten selbst genannt werden.
10c Kompetenzen erlangt	Aussagen darüber, wie die Evaluations-Kompetenzen in der Organisation erlangt worden sind.
11 Seminar	Aussagen darüber, welche Lerninhalte in einem Seminar über Evaluation gewünscht werden.
12 Evaluationsverständnis	Aussagen, die das Verständnis von Evaluation der Befragten widerspiegeln.



## **Erklärung**

Ich erkläre, dass ich die Dissertation unter Verwendung keiner anderen als der angegebenen Hilfsmittel angefertigt habe.

---

Ort, Datum, Unterschrift



